

Voice Conversion Challenge (VCC) 2016 participant questionnaire

(Please edit the space between questions to fill in your answers)

About your team:

1. Team name:
2. Affiliation:
3. How many people were involved in your entry for the VCC 2016?
4. How many person-hours were invested in total to prepare the material for your entry?

About your system and components:

5. If any, please specify the name of your system (e.g., Festival).
6. How long your system and/or the related research has been under development/study (e.g., 10 years)?
7. Is your system available commercially or as open source? If so, please specify it and/or provide a link with related online information.
8. If any, please specify components from common system(s) or toolkit(s) used in your system (e.g. Praat, STRAIGHT, etc.).

About the training process

9. Did you use data other than those provided for VCC 2016 to train the conversion pairs? If so, please specify the nature of the data (database) and how it was used in your system.
10. Did you use all the training data for this purpose or did you prune it?, If so please specify how you performed this selection.
11. Did you perform any manual task on the sentences of the training data (e.g., annotation, labeling, verification, etc.)?, If so, please specify what and the total number of person-hours spent on that task.
12. If any, please specify the system used for automatic phonetic labeling (e.g., HTK, Sphinx, etc.), please specify if it was used also when converting input (test) data.
13. If any, please specify the nature of the supra-segmental information used
14. Please specify the extracted features considered for training/estimation of the conversion models (e.g., LSFs, cepstrum, MFCCs, f0, energy, duration, etc.).
15. Please specify any pre-processing that was applied to the audio or features before training the conversion models.
16. Please specify and briefly describe (few lines) the alignment (if any) and the conversion model technique(s) (e.g. DTW, GMM, DNN, Freq. Warping, Linear Transformation, etc.) for ALL transformed features and the training/estimation approach used.
17. If any, please shortly describe the post-processing technique that was applied to improve/correct the features conversion or synthesis process.

18. Did any of points 9-17 have a pair-dependent characteristic? (Training variants according to the particular voices or gender of the speakers). If so, please specify clearly at which points and the nature of the dependence.
19. Please specify any other conversion or synthesis parameters individually set according to the conversion pair.

About signal processing and conversion process (synthesis)

20. Please specify the main techniques/models used for the acoustic analysis of the signal (harmonics, spectral envelope, pitch, energy, duration, etc.).
21. Is your system pitch synchronous? If so, please specify the pitch-marking technique.
22. Please clarify any differences between the analyses techniques/models used for the training and those used for the test data (e.g. different spectral features, harmonic model, etc.).
23. Please, shortly but clearly, describe the main transformation steps applied to ALL features considered by your system when converting an input utterance (few lines).
24. If any, please specify the excitation and/or phase prediction/correction technique for synthesis.
25. Please specify the technique, signal model or VOCODER used for transforming the input utterance after the features conversion process (if done separately).

About the computational cost

26. How many CPU hours did it take to build / train the model of a conversion pair?
27. What is the memory footprint of the system (conversion mode) at runtime?

Your opinions

1. Please share your general impression of the challenge.
2. Please provide your impression about the characteristics and quality of the databases of the VCC 2016.
3. Is your entry of the VCC 2016 representative of the best quality that can be achieved by your system? Please justify your answer (few lines).
4. Which component or contribution of your system for voice conversion purposes in general do you consider to be your strongest?
5. Which is the weaker component or task of your system?
6. Taking the performance of your system using the VCC data into account please tell us about your main motivation for future work.
7. Considering a next Voice Conversion Challenge:
 - a. Do you have suggestions regarding using database(s) and/or a modification of the rules?
 - b. Would you participate if a registration fee is required?
 - c. Would you be interested in volunteering to organize it?