Evaluation of a Multilingual Animated Speech System

Juliana De Ferran, Sonita Te & Stéphanie Monteiro
Under the supervision of researcher Slim Ouni

Motivation

In the field of audiovisual speech, animated virtual 3D faces of human speakers synced with audio (called talking heads) have been developed to model audiovisual speech communication and therefore study its mechanism.

Slim Ouni, Loria: 3 systems each animating the talking head's articulation of one language: English, French and German

VS

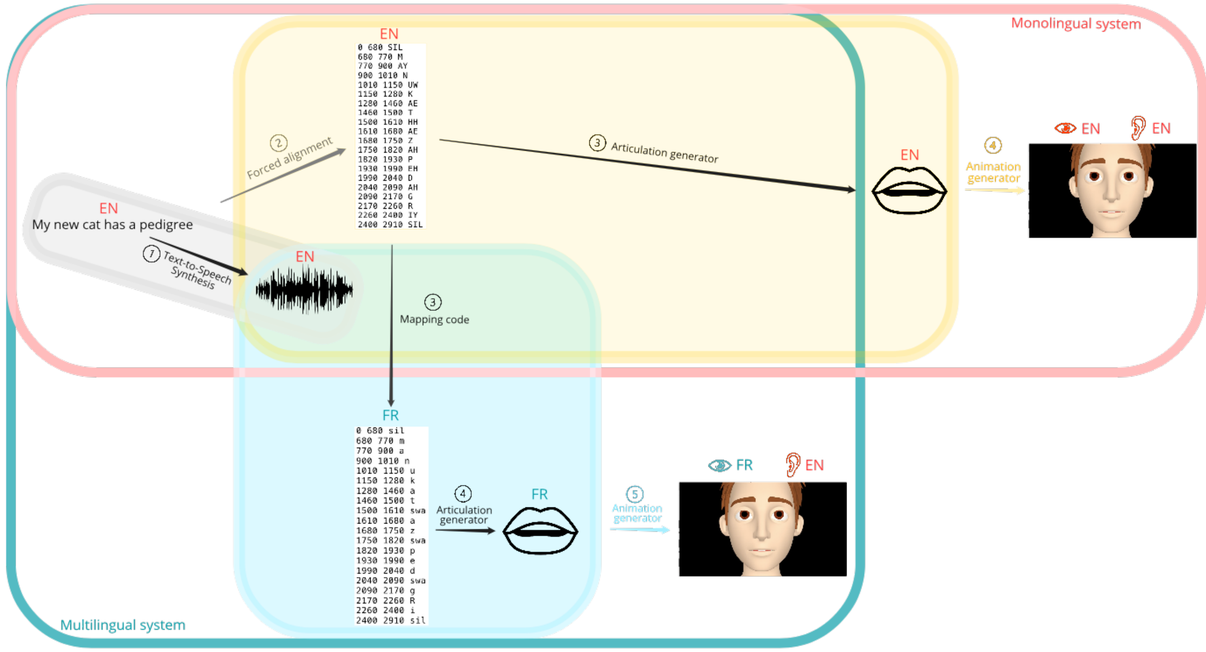
Taylor et al. : high quality audiovisual speech animation using a single system that animates any language

Previous work: study of the coarticulation, i.e. neighboring sounds that affect each other in the articulatory process, and intra/interlanguage differences → Hypothesis: monolingual system could be more accurate than multilingual system

Goal: create a multilingual system from what already exists and evaluate it using surveys.

Methodology

Diagram showing the functioning of the existing monolingual system and the multilingual system created



Creation of a multilingual system

➢ Existing system (red frame):

1. Text-to-Speech synthesis: create the audio corresponding to an input sentence in Language A
2. Forced alignment: generate a segmentation file where the sentence paired with the audio is sliced phone by phone and the beginning and end duration and the associated phone symbol are stored
3. Articulation generator: translate the segmentation file into articulation trajectories
4. Animation generator: generate a video from the audio in language A, the segmentation file in language A and the articulatory movements of language A, where what we hear is in Language A and what we also see.

➢ Our multilingual system (blue frame):

We use the same procedure, but instead of using the segmentation file of language A, we apply a mapping code (3) on it, which transforms each phone of Language A to each that most closely resembles them in Language B in the matter of articulatory characteristics. So, we get a segmentation file in language B. At the end, the system generates a video where the audio is in Language A but the articulation of the talking head is in Language B.

Evaluation

One survey per language, two parts in each survey

1. Rating independent videos
   - 30 random individual videos (10 mono, 10 multi Language B, 10 multi Language C)
   - scale: 1 to 5 (very bad to very good)
2. Comparison of videos
   - compare articulation of 3 videos (1st of the batch: from the monolingual system)
   - choose the closest to and furthest from the reality

Participants: natives, 10 minimum

Results

English survey

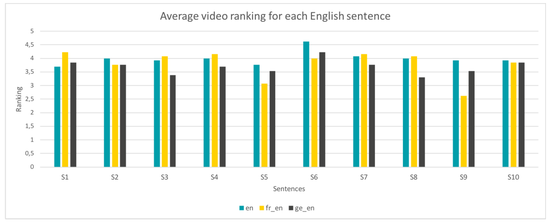
13 participants

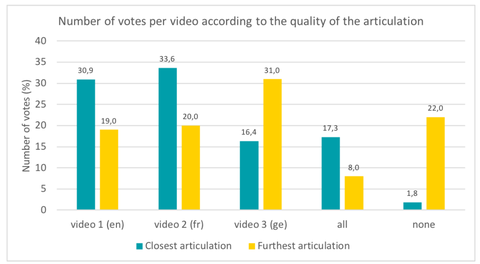
5/10 sentences: ranking average for articulation using both **monolingual and multilingual system**, especially the one mapping with French

33,6% of votes for video 2 (multi French) as containing the closest articulation
› 30,9% for video 1 (monolingual system)
› 16,4% for video 3 (multi German)

31% of votes for video 3 as containing the furthest articulation
› 20% for video 2
› 19% for video 1 (monolingual system)

→ Monolingual system and Multilingual system mapping with French





17 participants

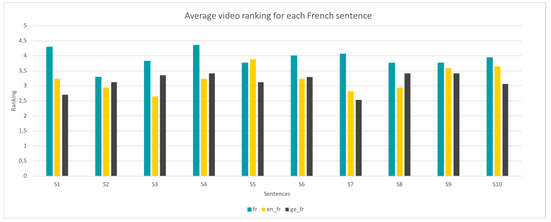
9/10 sentences : highest ranking average = articulation using **monolingual system**
› 1/10 sentences : articulation using the multilingual system mapping with English

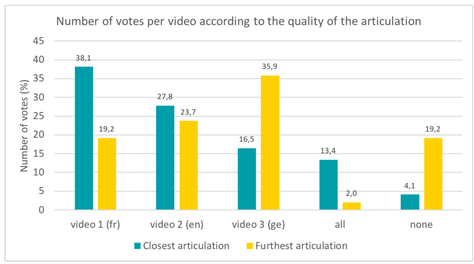
38,1 % of votes for video 1 (monolingual system) as containing the closest articulation
› 27,2% for video 2 (multi English)
› 16,5% for video 3 (multi German)

35,9% of votes for video 3 as containing the furthest articulation
› 23,7% for video 2
› 19,2% for video 1

→ Monolingual system

French survey





Conclusion

In view of the results, the articulation of monolingual system remains the better alternative. In addition, the quality of the animation of a multilingual speech system, while not terrible, would not allow this software to be used for things that call for extreme precision, such as lip rendering for hard-of-hearing individuals for example.

EN FR GE
MONOLINGUAL SYSTEM
Slim Ouni