

Формирование кадров при визуализации движения артикуляторных органов человека

Вяльцев В. Н.

Белорусский национальный технический университет

Визуальная информация очень важна при восприятии речи в шумах и незаменима для людей с ограниченным слухом или дефектами произношения. Существуют два подхода к проблеме визуализации движения артикуляторных органов человека при синтезе речи: имитационный и компиляционный.

В настоящий момент в ОИПИ НАН РБ занимаются разработкой системы имитационной визуализации движения артикуляторных органов человека. Существует законченная реализация 2D модели визуализации, основанной на системе синтеза речи, разработанной в Национальной Академии Наук Беларуси. В разработке также находится система 3D визуализации.

Формирование кадров в системе визуализации движения артикуляторных органов человека производится на основании так называемых визем – фотографий человека, произносящего различные звуки. Фотография разделяется на две части: неизменяемую – содержащую волосы, глаза, лоб, грудь и изменяемую – содержащую артикуляторные органы, важные для правильного восприятия визуальной речи. Неизменные части всех визем одинаковы. Они не перерисовываются при смене визем и не интерполируются. Изменяемые части всех визем различны. Они интерполируются при переходах. Всего для передачи всех звуков русского языка достаточно создать 20 визем.

В видеопотоке кадров виземы не могут сменять друг друга мгновенно, так как это будет грубо и неестественно. Поэтому при смене визем между ними создается область перехода. Область перехода формируется путем линейной интерполяции изменяемых частей двух соседних визем. Длительность перехода между виземами и смещение момента начала перехода от расчетного момента смены визем определяется длительностью отображения этих визем и рядом правил, определенных при анализе записи естественной речи.