

«Визуализация мимики лица диктора по его речи в условиях шумов»

Мазуренко И.Л., Холоденко А.Б.

dbabin@mech.math.msu.su

В последнее время увеличивается число фильмов, созданных на основе компьютерного графического моделирования различных объектов. Уже существует достаточное количество мультфильмов, созданных при непосредственном компьютерном участии, где двумерные или трехмерные анимированные компьютером персонажи двигаются и разговаривают как живые. Однако, положение губ и лица говорящего героя обычно формируется заранее, программно, и, следовательно, не позволяет создавать анимированные фильмы «на лету». Хотелось бы, чтобы мимика персонажей при разговоре формировалась в реальном времени по движению губ человека, надиктовывающего эту фразу в микрофон.

Актуально создание кукол с автоматической мимикой. Использование метода синтеза движения губ по звуковому сигналу дает значительный выигрыш по сравнению с существующими ручными методами. Существуют подобные системы, ориентированные на английскую/американскую речь, и, при этом, стоящие значительную сумму денег.

Актуальна задача создания компьютерного пакета, который мог бы демонстрировать работу алгоритмов мимико-речевого соответствия независимо от языка говорящего.

Работа существующих подобных системы (voiceDSP фирмы Ventriloquist и система Chandrika Krothapalli, David F. McAllister, Robert D. Rodman, Donald Bitzer, Meng Wang and Jamie Taylor) зависит от языка (только английский), диктора, и шумовой обстановки.

Работающий прототип русскоязычной системы, преобразующей речевой сигнал в параметры движения губ, был разработан в нашем коллективе в 2002-2003 годах. Для этого использовались метод скрытых Марковских процессов, русский фонетический корпус, многомикрофонная система распознавания фонем в шумах.

Нами был выбран алфавит из 8 русских визем (пауза, А, О, У, И, Э, С, М) и построено многомерное кусочно-линейное преобразование, позволяющее спектральные параметры вводимой с микрофона речи преобразовывать в вероятностную последовательность «визем». Более точно, для каждого окна вводимой в компьютер с микрофона речи длительностью около 20 мс рассчитывались спектральные параметры, на основе которых с помощью некоторого преобразования получался вектор вероятностей наблюдения каждой из визем в данный момент времени. Кусочно-

линейное преобразование получалось путем обучения на примерах из специально созданных фонетических баз данных с видеоинформацией. Полученные с помощью этого преобразования вероятностные распределения визем корректировались с учетом непрерывности изменения параметров движения губ, а затем преобразовывались к значениям ширины и высоты рта анимированного персонажа.

Список публикаций

1. А.Б. Холоденко. Исправление ошибок в формальных языках
Материалы конференции "Нейрокомпьютеры и их применение". Москва, 2000 г.
2. И. Л. Мазуренко. Многоканальная система распознавания речи
В сб.: VI всероссийская конференция "Нейрокомпьютеры и их применение".
Сборник докладов. Москва 16-18 февраля 2000 г. Стр. 222-225.
3. Мазуренко И.Л. Компьютерные системы распознавания речи.
Интеллектуальные системы, т.3. вып. 1-2 – Москва, 1998 г. Стр. 117-134.
4. Мазуренко И.Л. О сокращении перебора в словаре речевых команд в составе
системы распознавания речи. Интеллектуальные системы, т.2, вып. 1-4, Москва,
1997. Стр. 135 – 148.
5. Д.Н.Бабин, И.Л.Мазуренко, А.Б. Холоденко., О перспективах создания системы
автоматического распознавания слитной устной русской речи, Интеллектуальные
системы, т.6, вып. 1-4, Москва, 2003. Стр. 5 – 24 .
6. <http://intsys.msu.ru/invest/speech/articles/>