

Международный научно-учебный центр информационных  
технологий и систем

Национальный политехнический университет Украины  
«Киевский политехнический институт»  
Факультет Электроники

# АННОТАЦИЯ И УЧЕТ РЕЧЕВЫХ СБОЕВ В ЗАДАЧЕ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ СПОНТАННОЙ УКРАИНСКОЙ РЕЧИ

**В. В. Пилипенко, О. Н. Ладоско**

# Содержание

- **Введение**
- **Система распознавания слитной речи**
  - Программный комплекс
  - Особенности речи дикторов
- **Речевые сбои спонтанной речи**
  - Причины возникновения
  - Классификация
- **Основные типы речевых сбоев**
  - Заполненные паузы
  - Самоисправления
  - Неверно произнесенные слова
  - Влияние на процесс распознавания речи
- **Акустическое и текстовое наполнение**
  - Обучающая и анализируемая выборки
  - Текстовый материал
- **Анализ речевых сбоев группы дикторов**
  - Особенности появления речевых сбоев
- **Результаты экспериментов**
  - Анализируемой выборки
  - Контрольной выборки
- **Заключение**

# Введение

## Цель работы:

Исследование влияния характерной особенности спонтанной речи – речевых сбоев, на показатели надёжности системы автоматического распознавания речи

## Актуальность:

- Речевые сбои **неформальной** разговорной речи **15-20%** всех от слов.
- **Усложнение** процесса автоматического распознавания спонтанной речи

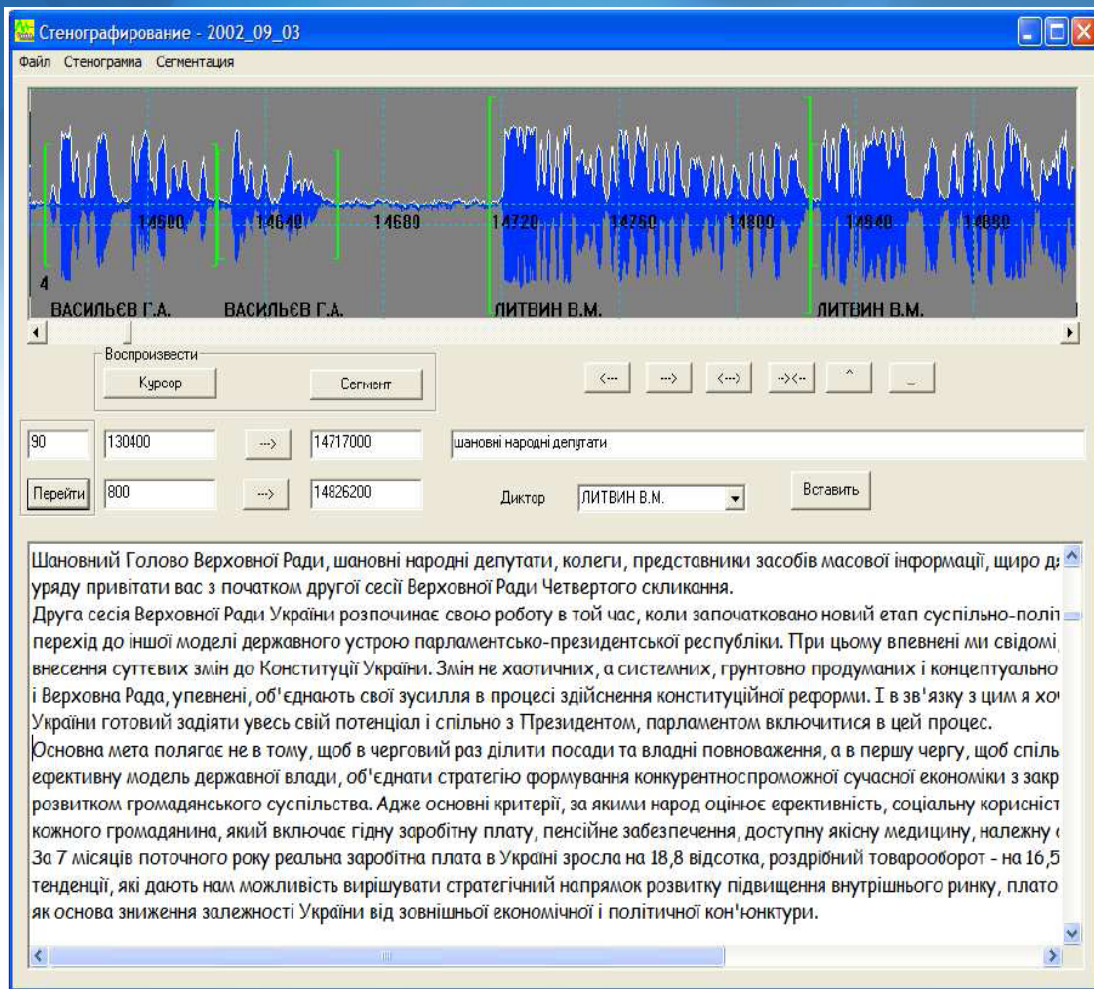
## Применение:

- Системы стенографирования речи: получение текста стенограммы из звукового файла
- Обработка материалов различных заседаний, форумов...

## Как решалась задача:

- Учет речевых сбоев
- Коррекция стенограммы
- Очистка речевых данных
- Улучшение показателей надёжности системы распознавания речи.

# Система распознавания слитной речи



Пользовательский вид программы стенографирования

## Особенности речи дикторов:

- Спонтанная речь
- Частые отступления от подготовленного доклада
- Ограничение во времени приводит к повешению темпа речи
- Эмоционально окрашена
- Некоторые голоса обладают тембральной немелодичностью
- Вмешательство ведущего заседания и других депутатов
- Высокое качество записи, каждое место снабжено микрофоном

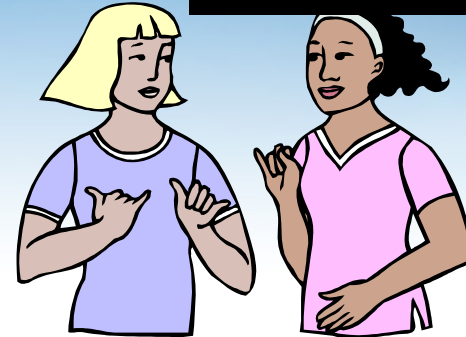
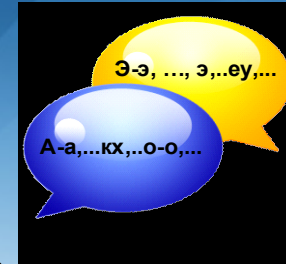
# Речевые сбои спонтанной речи

## Речевые сбои:

- Неинформативные элементы (колебания, повторы, исправления)
- Помеха, искажающая исходный языковой материал

## Могут быть вызваны:

- Самим диктором (придыхания, откашливания)
- Нарушением внешних условий коммуникации (вмешательство собеседника)
- Внутренними проблемами планирования дискурса



## Основные причины появления

- Ускоренный темп речи
- Фонационное дыхание
- Не соблюдение орфоэпических норм



# Речевые сбои спонтанной речи



# Основные типы речевых сбоев

## Заполненные паузы

### Одинарного звучания



|(е) привернуть |(е) цю |(е) сферу

### Нефонологическая долгота



Наших громадян було в |(е) визволено

### Откашливание



В мене питання. Чи ви |(кг)

## Виды:

- заполненные паузы, напоминающие фонемы «а», «е», «о»
- растягивание звуков «ее», «ме», «аа» («эkanie», «мэkanie», «акание»)
- комбинации звуков «хе», «кг», «ги» (придыхания, откашливания)

## Функции:

- заполнителей промежутков в речи
- обдумывания и перепланирования речи
- позволяют избежать разрыва фраз
- вставок в самоисправления диктора

**Заполненные паузы** – это, так называемые, вокализованные паузы.

# Основные типы речевых сбоев

## Самоисправления

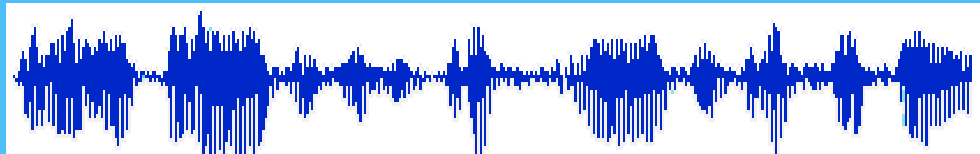
### Фальстарты слов



процвітаючою, великою, |(поту) потужною державою

- **Самоисправления** - коррекции направленные на уточнение сказанного
- “Пробные шаги” диктора

### Коррекции со вставкой



Шановні колеги! Я хочу |(попрос) |(е) |попросити,

- Использование вставок позволяет выделить большее время для осмысления варианта исправления

### Он-лайн коррекция



Коли у видатках |(бурже) |бюджету

- Забракованный фрагмент речи полностью отличается от откорректированного варианта



# Основные типы речевых сбоев

## Самоисправления

### Обрыв слов



наполягаєте ви на цій |(поправ...)

В случае **обрыва** диктор не договаривает начатое слово:

- без последующей коррекции (индивидуальное произношение)
- планирует дальнейшее исправление: (фальстарты и коррекции)
- Обрывы могут быть частью других видов коррекций

### Обрыв слов в повторе



платник податку або не платник |(пода...)

# Самоисправления

## Самоисправления

### Повторы слов



|зауваження

|зауваження стосувалися

### Повторы слов со вставкой



до залу,

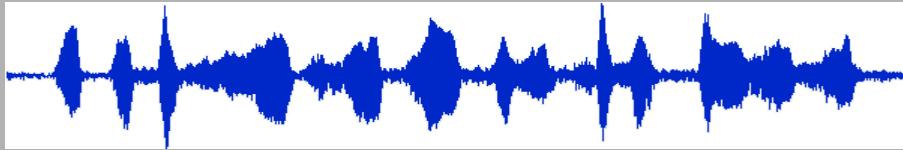
|(який я) |за

|(який я) доповідав

- Повторы слов совмещают функции колебания и коррекции
- Повторы выражают сомнения диктора в выборе подходящего фрагмента:
  - по форме фрагмента
  - по значению фрагмента
  - по месту положения фрагмента
  - фрагмента находящегося в процессе обдумывания
- Вставка позволяет выделить большее время для переосмысления сказанного фрагмента

# Неверно произнесенные слова

## Слова произнесенные неверно

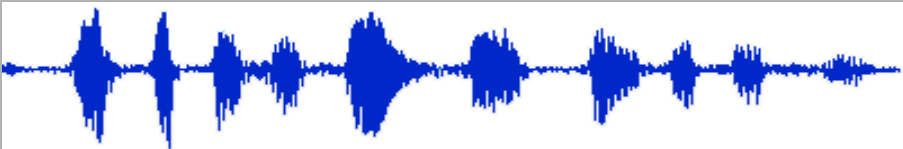


(епідомічного) благополуччя населення

- Характерны черты редуцирования (изменения) нескольких звуков

епідемічного → епідоммічного

## Беспорядочные слова



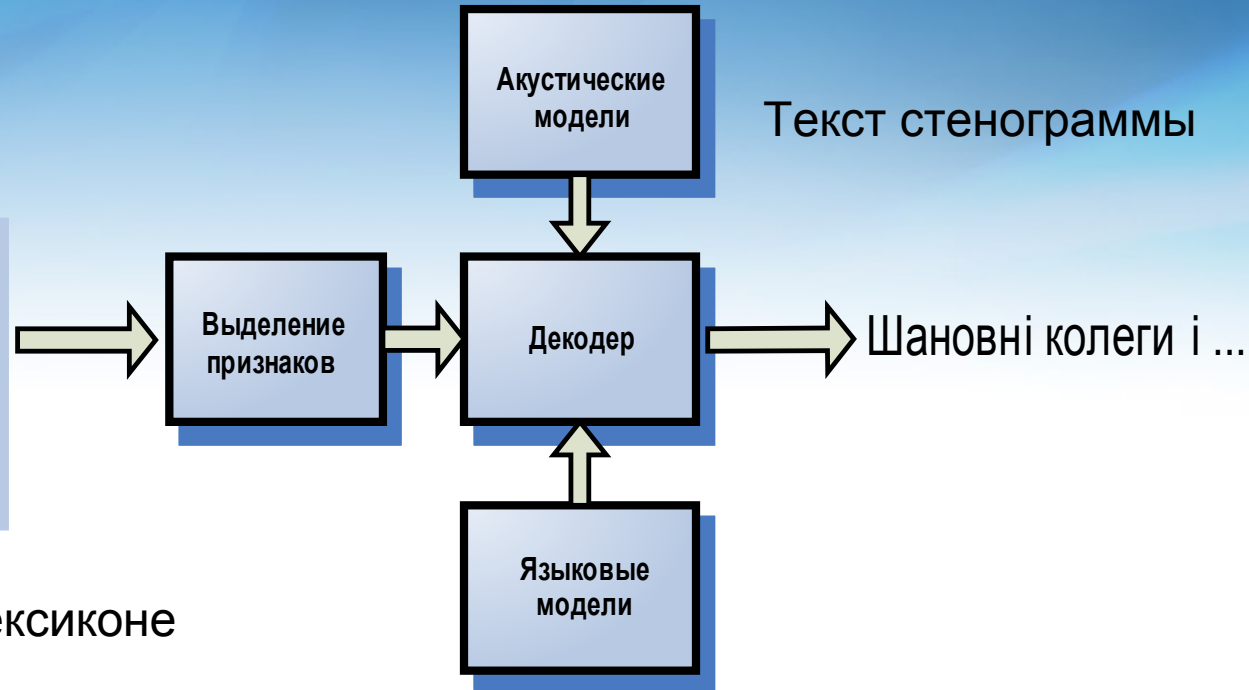
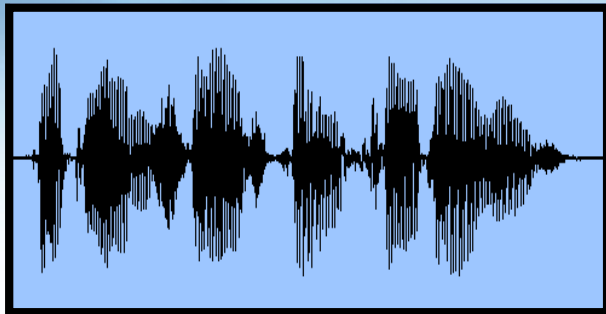
двісті ((під)тисізти) тисяч осіб

- Смысл беспорядочных слов обнаруживается только из контекста речи
- Причины возникновения беспорядочных слов:
  - Ускоренный темп
  - Не соблюдение орфоэпических норм

**Беспорядочные слова** – это слова, в которых нарушена правильная последовательность появления фонем

# Влияние речевых сбоев на процесс автоматического распознавания речи

Звуковое сопровождение



- Нет представления в лексиконе
- Нет представления новых звуков
- Не полное соответствие языковой модели
- Генерирование неверной гипотезы распознавания слова

Один речевой сбой может быть причиной появления более чем одной ошибки

# Акустическое и текстовое наполнение

## Текстовый материал:

- Словарь из текстов стенограмм заседаний ВРУ – более **100МБ** текста (**14М** слов)

## Анализируемая выборка:

- **11** файлов заседаний ВРУ
- длительность **35390 сек (10 часов)**
- встретилось **73 775 слов**
- 203 докладчика
- дикторы, которые произнесли **500** и более слов **46**
- 8 дикторов из 46 выступали более одного раза

## Обучающая выборка:

- Записи длиной **197 762 сек (55 часов)**
- Встретилось **427 576 слов**
- записано **287 дикторов**
- **139 дикторов** с длительностью речи каждого более **300 сек**

## Контрольная выборка:

- **7 файлов** записей ВРУ
- длительность **36949 сек (10 часов)**

# Анализ характерных речевых сбоев группы дикторов

## Обработка текстового материала

1. Введена ручная разметка речевых данных для выявления неинформативных элементов, обнаружения условий и контекста их появления.
2. Выполнена коррекция текста стенограммы
3. Внесены недостающие элементы спонтанной речи
4. Подсчет расстройств речи конкретных дикторов и общей выборки дикторов реализован автоматически
5. Проведено автоматическое удаление сегментов речи, которые содержат любой из вышеуказанных типов сбоев

## Проблемы выявления речевых сбоев

- Не имеют лексического представления (не принадлежат к словам данного языка)
- Не имеют грамматического представления (нет грамматических правил записи)
- По-разному описываются в зависимости от типа речевого сбоя





# Анализ характерных речевых сбоев группы дикторов

**Анализируемая выборка дикторов (более 500 слов и наибольшее число сбоев)**

Особенности диктора	GRY	ZVA	USH	TER	CIB	SAH	Сумма сбоев	сумма сбоев/общ. число сбоев, %
Заполненная пауза «а»	4	12	4	47	22	25	114	40,86
Заполненная пауза «ее»	1	0	0	7	7	7	22	7,89
Фальстарты слов	0	1	1	2	4	0	8	2,87
Повторы	1	0	4	8	1	1	15	5,38
Обрывы	0	1	0	7	0	1	9	3,23
Коррекция со вставкой	0	1	3	4	4	1	13	4,66
Он-лайн коррекция	2	5	4	9	10	3	33	11,83
Повторы со вставкой	0	1	1	3	0	0	5	1,79
Слова произнесенные неверно	20	3	14	12	6	1	56	20,07
Беспорядочные слова	0	0	1	1	1	1	4	1,43
Общ. кол. ошибок	28	24	32	100	55	40	279	100

- Заполненные паузы «а» **40,86%**
- Слова произнесенные неверно **20,07%**
- Он-лайн коррекция и коррекция со вставкой **16,49%**
- Заполненные паузы «ее» **7,89%**
- Остальные не превышают **5,38%**

Общее число речевых сбоев: **6 дикторов - 279 шт.**

# Анализ характерных речевых сбоев группы дикторов

Дикторы, которые произнесли 500 и более слов и совершили наибольшее число речевых сбоев

Особенност и диктора	GRY	ZVA	USH	TER	CIB	SAH	Сумма
Общ. кол. слов	660	513	663	1865	979	576	5256
Общ. кол. ошибок	28	24	32	100	55	40	279
Общ. кол. ошибок в докладе, %	4,24	4,68	4,83	5,36	5,62	6,94	5,31
Время выступлен ия, сек	354,88	252,06	305,79	874,05	448,30	287,03	2522,11
Темп, слов/сек	1,86	2,04	2,17	2,13	2,18	2,01	-

- Увеличение количества ошибок в зависимости от повышения темпа речи.
- С увеличением количества слов доклада частота появления речевых сбоев возрастает.
- Речевые сбои отдельно взятого диктора распределены неравномерно.

Практика показала, что содержание речевых сбоев **существенно зависит от ораторских способностей диктора**, поэтому **содержание речевых сбоев** речи следует исследовать для **каждого диктора в отдельности**

# Результаты экспериментов

## Речевые сбои выборки 11 файлов

Файл	«а»	«ее»	Фальс-тарт	Повтор	Обрыв	Корр.	Слова произн. неверно	Беспор. слова	Общ.ко л. сбоев
10_16	27	15	3	3	1	11	27	5	92
09_26	29	24	4	10	8	16	40	1	132
10_08	42	17	2	8	9	41	108	12	239
10_18	43	11	5	8	9	28	73	7	184
10_22	69	29	6	9	7	37	63	5	225
10_23	55	13	9	7	1	16	14	3	118
10_24	99	13	10	36	35	49	92	6	340
10_25	63	16	6	11	7	15	20	5	143
11_12	16	6	7	3	18	18	29	9	107
11_19	35	5	2	2	1	6	11	1	63
11_20	45	21	6	8	4	30	38	8	160
<b>Сумма по выборке</b>	<b>523</b>	<b>170</b>	<b>60</b>	<b>105</b>	<b>100</b>	<b>267</b>	<b>516</b>	<b>62</b>	<b>1803</b>

### Анализируемая выборка:

- **11 файлов** заседаний ВРУ
- длительность выступлений **35390 сек**
- **73 775 слова**
- **203 докладчика**
- Дикторы, которые произнесли **500** и более слов **46**
- **8 дикторов** из **46** выступали более одного раза

- **Общее число речевых сбоев**, отмеченное на выборке 11 файлов, составляет **1803 единиц – 2,44%**
- Заполненные паузы одинарного звучания «а» **523 шт. - 28,95%**
- Слова произнесенные неверно **516 шт. - 28,62%**
- Коррекции и фальстарты слов **327 шт. - 18,14%**
- Заполненные паузы с растягиванием звуков «ее» **170 шт. - 9,43%**
- **Остальные типы сбоев не превышают 105 шт. - 5,82%.**

# Результаты экспериментов

Файл	Общ. кол. сбоев	Общ. кол. слов	Сбои/общ. кол. слов, %	Время выступ.	Средний темп, слов/сек	Надёжн. распозн. реч.сбои, %	Надёжн. распозн. очищ.реч., %	Изменение, %
10_16	92	5403	1,70	2512	2,19	76,07	79,61	3,54
09_26	132	2991	4,41	1369	2,19	82,94	84,26	1,32
10_08	239	7007	3,41	3354	2,11	78,55	81,72	3,17
10_18	184	7485	2,46	3394	2,19	69,93	72,9	2,97
10_22	225	11525	1,95	5978	1,89	73,71	73,95	0,24
10_23	118	4595	2,57	2205	2,03	74,62	74,95	0,33
10_24	340	9257	3,67	4114	2,17	76,28	77,16	0,88
10_25	143	5942	2,40	2794	2,17	63,83	64,02	0,19
11_12	107	8208	1,30	4280	1,94	70,82	71,19	0,37
11_19	63	4311	1,46	2004	2,18	74,99	75,44	0,45
11_20	160	7051	2,27	3386	2,09	73,89	74,17	0,28
Сумм. сбоев	1803	73775	2,44	35390	-	-	-	1,25

## Речевые сбои:

- **1803 шт. - 2,44%**
- 46 дикторов
- 73 775 слова
- длительность выступлений 35390 сек
- Повышение надёжности в среднем на **1,25 %**
- Максимальное повышение надёжности **3,54 %**

**Все виды аннотируемых речевых сбоев оказывают негативное влияние на систему распознавания речи.**

# Результаты экспериментов

## Результаты распознавания контрольной выборки

Файл	Длина, сек	Надёжн. распозн. сбой, %	Надёжн. распозн. очищ. речь, %	Изменение, %
09_10a	4925	66,22	71,46	5,24
09_27	1521	74,40	81,19	6,79
10_15	8678	70,56	77,30	6,74
10_15a	5022	71,29	78,85	7,56
10_17	6088	79,09	84,37	5,28
10_17a	3282	76,65	82,60	5,95
11_21	7433	75,79	83,41	7,62

## Контрольная выборка:

- 7 файлов записей заседаний
- длительность выступлений **36949 сек**
- на **контрольной выборке** не производилось обучение системы распознавания речи
- произведена **коррекция** и **проверка** грамматики стенограммы
- проведена **разметка** речевых данных, включающих **элементы спонтанной речи** депутатов
- Детальный **анализ** речевых сбоев **не проводился**

Результат распознавания на **контрольной выборке** дал повышение надёжности распознавания в среднем на **6,45%**

# Заключение

- Выявлены и систематизированы основные типы речевых сбоев в спонтанной украинской речи
- Анализ речевых сбоев группы дикторов показал, что содержание речевых сбоев существенно зависит от ораторских способностей диктора, поэтому содержание речевых сбоев речи следует исследовать для каждого диктора в отдельности
- Обработка речевых данных позволила выявить **1803 речевых сбоя** для **анализируемой выборки** и улучшить показатели надёжности распознавания речи для **анализируемой** и **контрольной выборки** соответственно в среднем на **1,25 %** и **6,45%** соответственно.

**Полученные данные и результаты могут быть в дальнейшем использованы для:**

- Обработки нефонемных элементов на акустическом уровне
- Разработки подходов учета структуры нефонемных элементов
- Моделирования элементов присущих спонтанной украинской речи



# Заключение

Представленные результаты предварительного исследования подтвердили существенность влияния неинформативных элементов спонтанной речи на показатели надёжности системы автоматического распознавания слитной речи.

**Полученные данные и результаты будут в дальнейшем использованы для:**

- Обработки нефонемных элементов на акустическом уровне
- Разработки подходов учета структуры нефонемных элементов
- Моделирования элементов присущих спонтанной украинской речи

**Описанные подходы направлены на:**

- Исследования нефонемных элементов спонтанной речи
- Автоматизацию процесса выявления и разметки

**Автоматизирование процесса выявления и разметки речевых сбоев позволит уменьшить долю ручного аннотирования и охватить больший круг описываемых явлений.**

# Спасибо за внимание!

