

Розпізнавання мовленнєвого сигналу

Микола Сажок

*Відділ розпізнавання та синтезу звукових образів,
Міжнародний науково-навчальний центр
інформаційних технологій та систем*



2022

Світова спадщина - Legasy

Audrey (1952), Bell Labs,



Shoebox (1961), IBM



Цифри, 0-9

DARPA SUR (1971–1976)

Мета: 1000 слів, 90%

Harpy, CMU

Hearsay II, CMU

HWIM, BBN



Tangora (1980s), IBM —
20 000 слів



James Flanagan, Lawrence Rabiner, Raj Reddy, Fred Jelinek, H. Fujisaki, H. Hermansky, S. Furui, Hermann Ney, Jerard Chollet

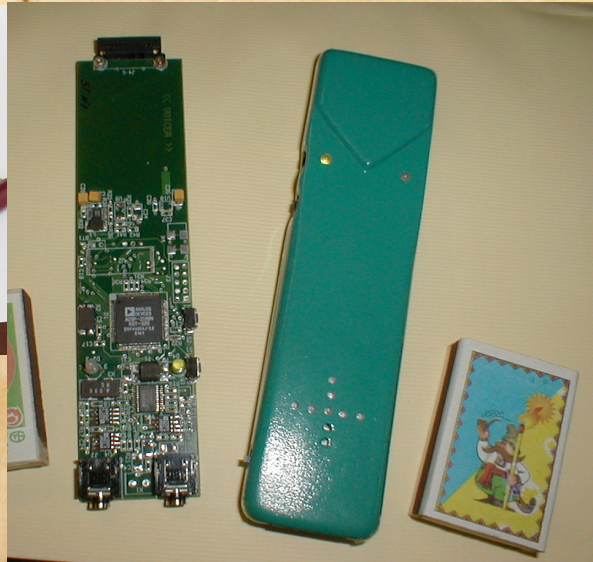
Вітчизняна Спадщина

Пофонемне розпізнавання,
злите мовлення тривалістю до 25 с,
пофонемний синтез мовлення за текстом,
7 мов, розроблена на замовлення ЮНЕСКО

Проф. Тарас Вінцюк демонструє
експериментальну систему розпізнавання
зв'язного мовлення на ЕОМ "БЭСМ-6" (1970)

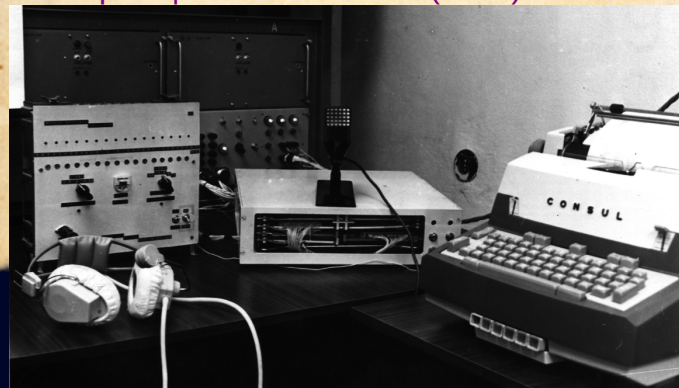


Вокофони – портативні цифрові
диктофони, керовані голосом (2002, 2004)

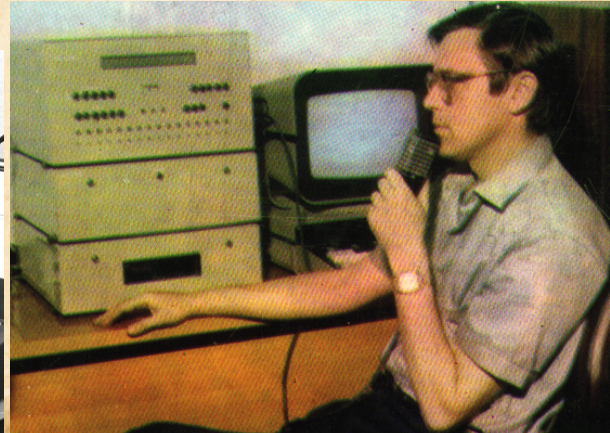


Система усного діалогу Речь-121
(1986)

Пристрій розпізнавання усних команд на
Електроніці-60 і ... "Consul" (1979)

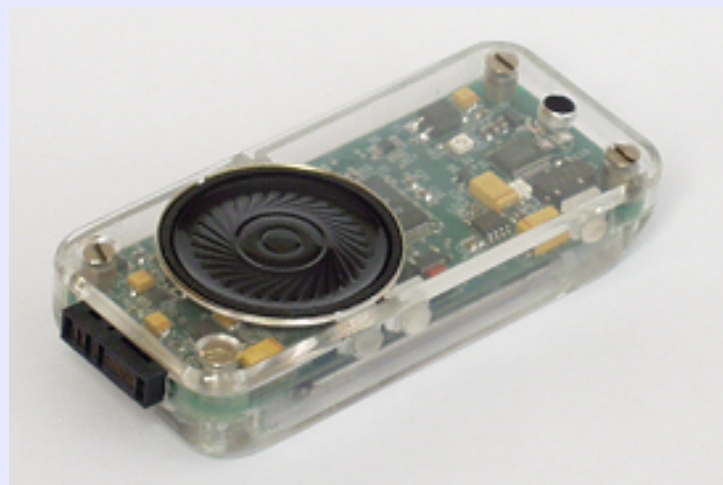


Система усного діалогу Речь-1 (1981)



ДНТП “Образний комп’ютер”

2000 – 2006: Вокофони – портативні цифрові диктофони з голосовим управлінням



ДНТП “Образний комп’ютер”

Проф. Вінцюк презентує Вокофон в Українському домі - 2002



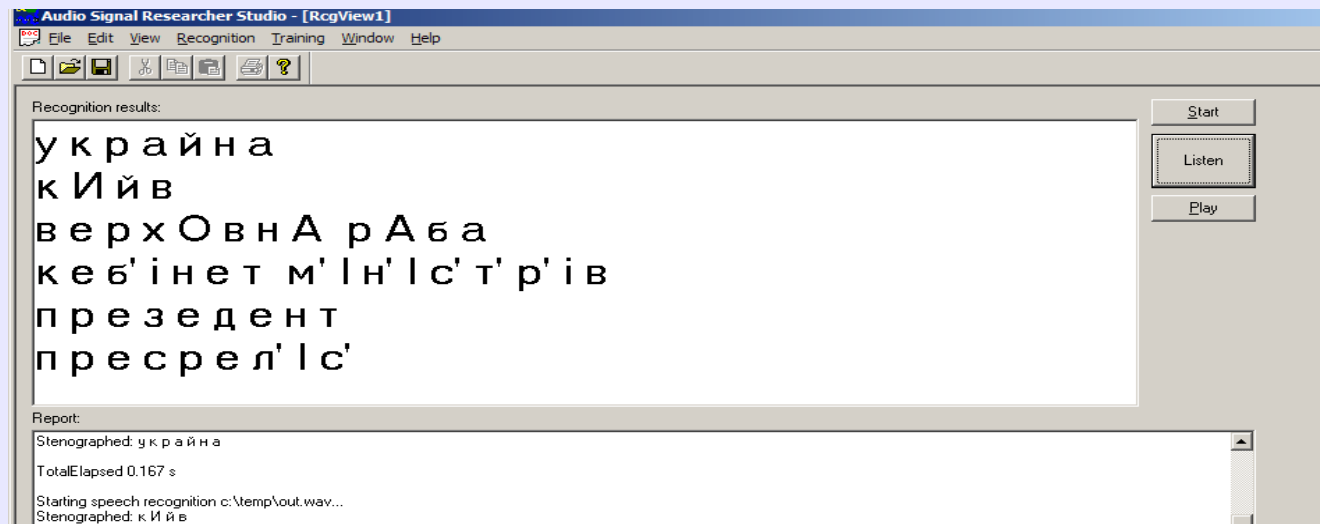
Практичні результати та ступінь їх готовності до передачі у виробництво, основні техніко-економічні характеристики

Система швидкого автоматичного розпізнавання слів з великих словників для ПК на основі фонетичного стенографа.

Є експериментальні зразки.

Українська
мова

55 фонем

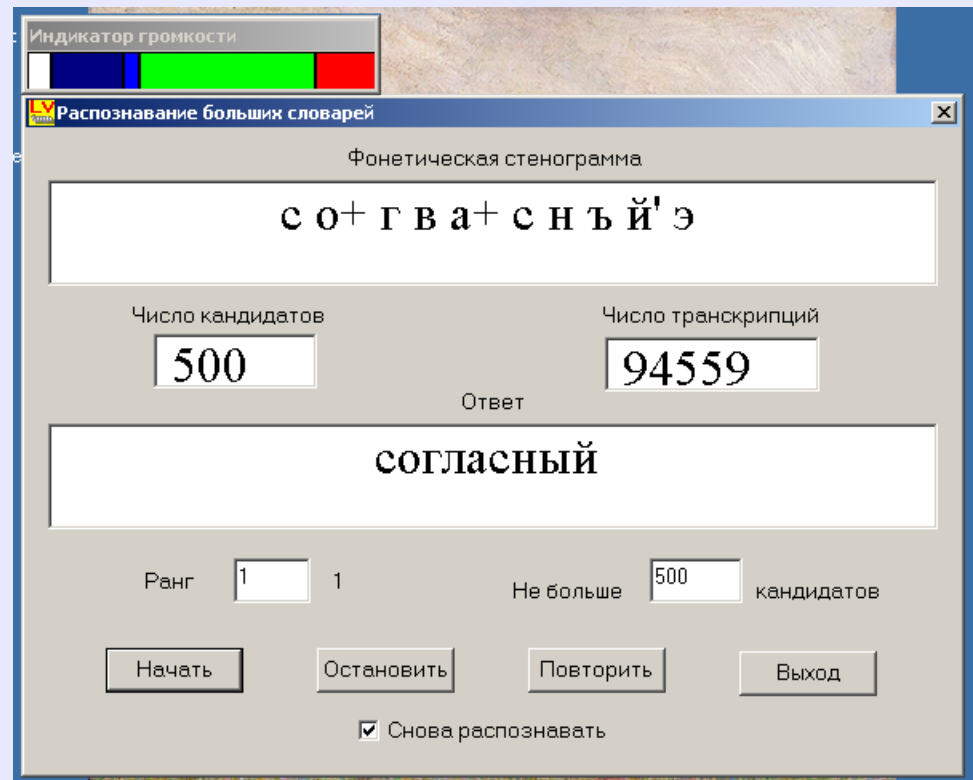


Практичні результати та ступінь їх готовності до передачі у виробництво, основні техніко-економічні характеристики

Система швидкого автоматичного розпізнавання слів з великих словників для ПК на основі фонетичного стенографа. Є експериментальні зразки.

Російська мова

Система працює зі словниками до 95 тис. слів у реальному часі. Надійність 96%. Є варіанти з робочим словником до 1 млн. слів.



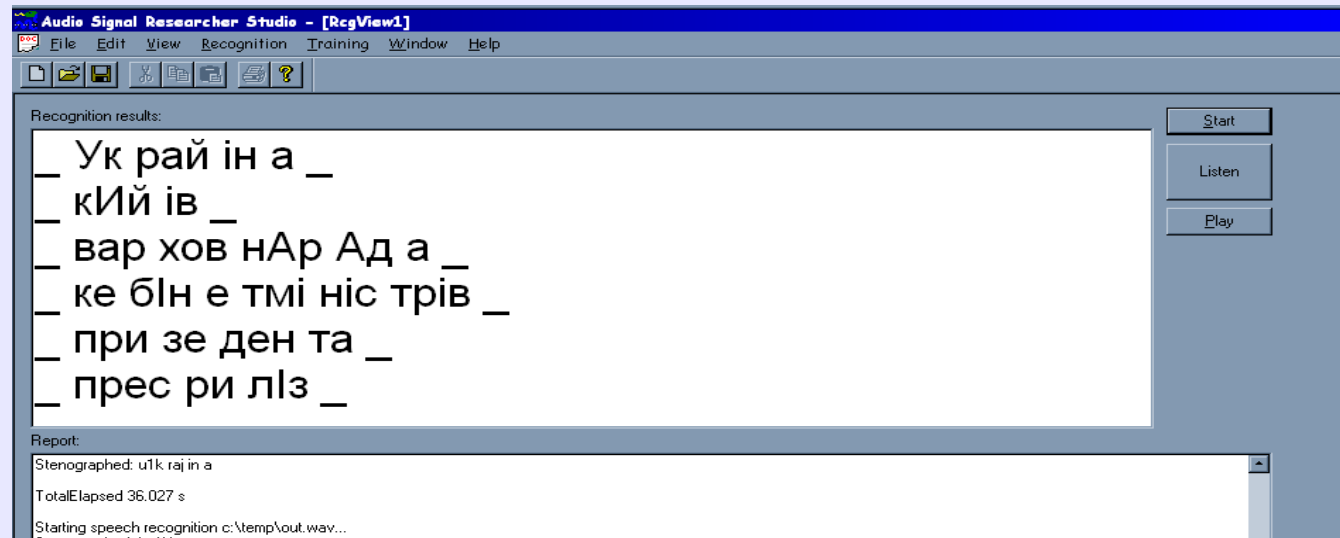
Практичні результати та ступінь їх готовності до передачі у виробництво, основні техніко-економічні характеристики

Система швидкого автоматичного розпізнавання слів з необмежених словників для ПК на основі поскладового стенографа.

Є експериментальні зразки.

Українська
мова

10000
складів



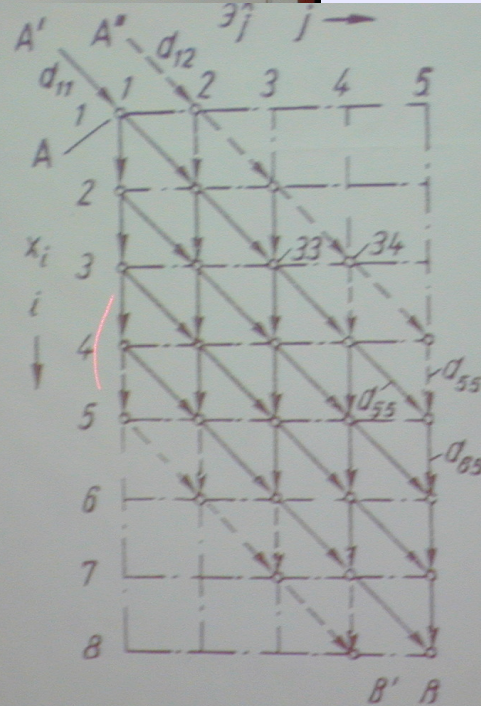
Speech recognition research in 1950s and 1960s (2)

DTW

- Elementary time-normalization methods by Martin et al.
- Vintsyuk in Soviet Union proposed the use of DP
- Sakoe & Chiba at NEC

Continuous speech re

- Reddy at CMU conduc
- based on dynamic trac



Dynamic time warping algorithm
(T. K. Vintsyuk, Lab of Pattern Recognition, USSR, 1968)



ECSH 2021 BRNO CZECHIA

Bayes Decision Theory and Machine Learning

RWTH AACHEN
UNIVERSITY

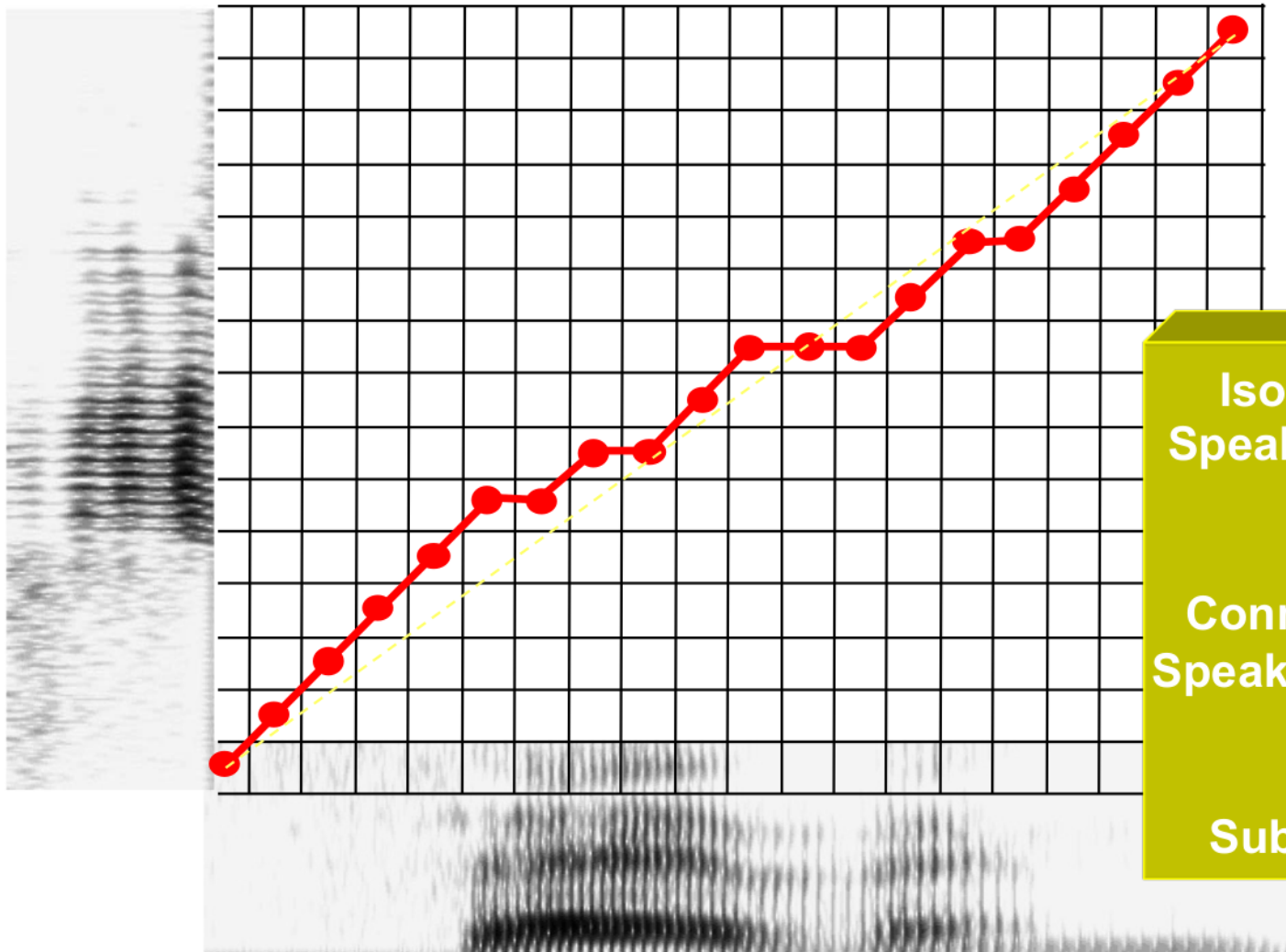
principal ingredients:

- performance measure, error measure, loss function:
 - how to judge the quality of the system output
 - examples: ASR: edit distance; MT: TER or BLEU
- probabilistic models (with a suitable structure)
for capturing the dependencies within and between input and output strings:
 - given synchronization: Markov chain, CRF, (LSTM) RNN, ...
 - input/output synchronization: generative/hybrid HMM, CTC, RNN-T, attention models, ...
- training criterion:
 - to learn the free model parameters from examples
 - ideally should be linked to performance criterion
 - two open questions: exact form of criterion? optimization strategy?

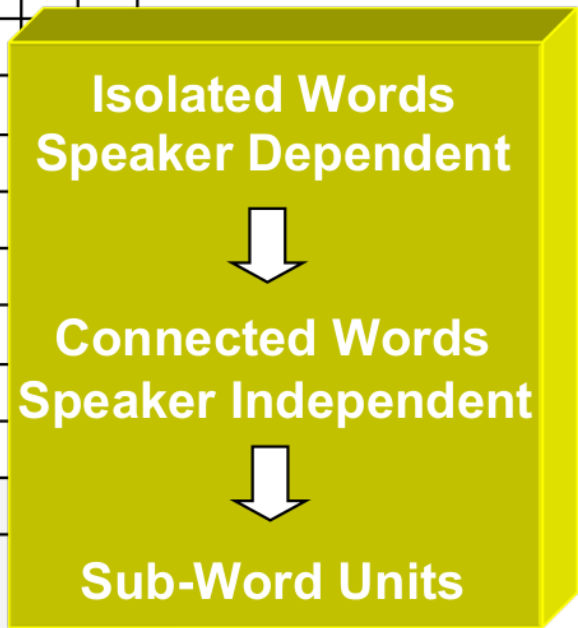
- Bayes decision rule: decoder/search
for generating the output word sequence
 - combinatorial problem (efficient algorithms)
 - should exploit structure of models
 - examples: dynamic programming and beam search, A* and heuristic search, ...
[Vintsyuk 68, Velichko & Zagoruyko 70, Vintsyuk 71] and [Sakoe & Chiba 71],
[DRAGON/HARPY 1975] and [Bridle 82, Ney 84, Ney & Haeb-Umbach 92]

Dynamic Time Warping (DTW)

TEMPLATE (WORD 7)



T.K. Vyntsyuk (1968)
H. Sakoe,
S. Chiba (1970)



UNKNOWN WORD

Актуальність

- перетворення мовлення на текст (make speech searchable!);
- технології розпізнавання мовлення для аналізу телерадіоефіру включно з інтернет-джерелами;
- розпізнавання особи за голосом (діаризація, ідентифікація, верифікація);
- аналіз тональності мовлення (sentiment analysis);
- структурування мітингів (summarization, action items);
- усний переклад відеороликів;
- автоматизація кол- і контакт-центрів;
- аналіз мовленнєвої комунікації співробітників;
- ?

- ❖ Системи усного диктування, редагування, голосова клавіатура;
- ❖ підвищення стійкості до шумів, завад;
- ❖ робота з телефонними, ефірними, медійними каналами;
- ❖ моделювання індивідуальних особливостей мовця;
- ❖ багатомовність;
- ❖ розробка діалогових систем, смислова інтерпретація, мультимодальність;
- ❖ подальше підвищення споживчих якостей розроблюваних мовленнєвих інформаційних технологій та електронних пристроїв.

Ієрархія мовленнєвих образів

Акустичний рівень

- Мікрофонема: вікно аналізу, сегмент квазіперіоду
- Фаза фонемі, сенон
- Базова фонема: контекстно-незалежні (монофони), контекстно-залежні (фонемі-трифони)

■ Фонемосполучення: дифони, склади, морфемі

■ Слово

Синтаксичний рівень

■ Словосполучення, ритмогрупа

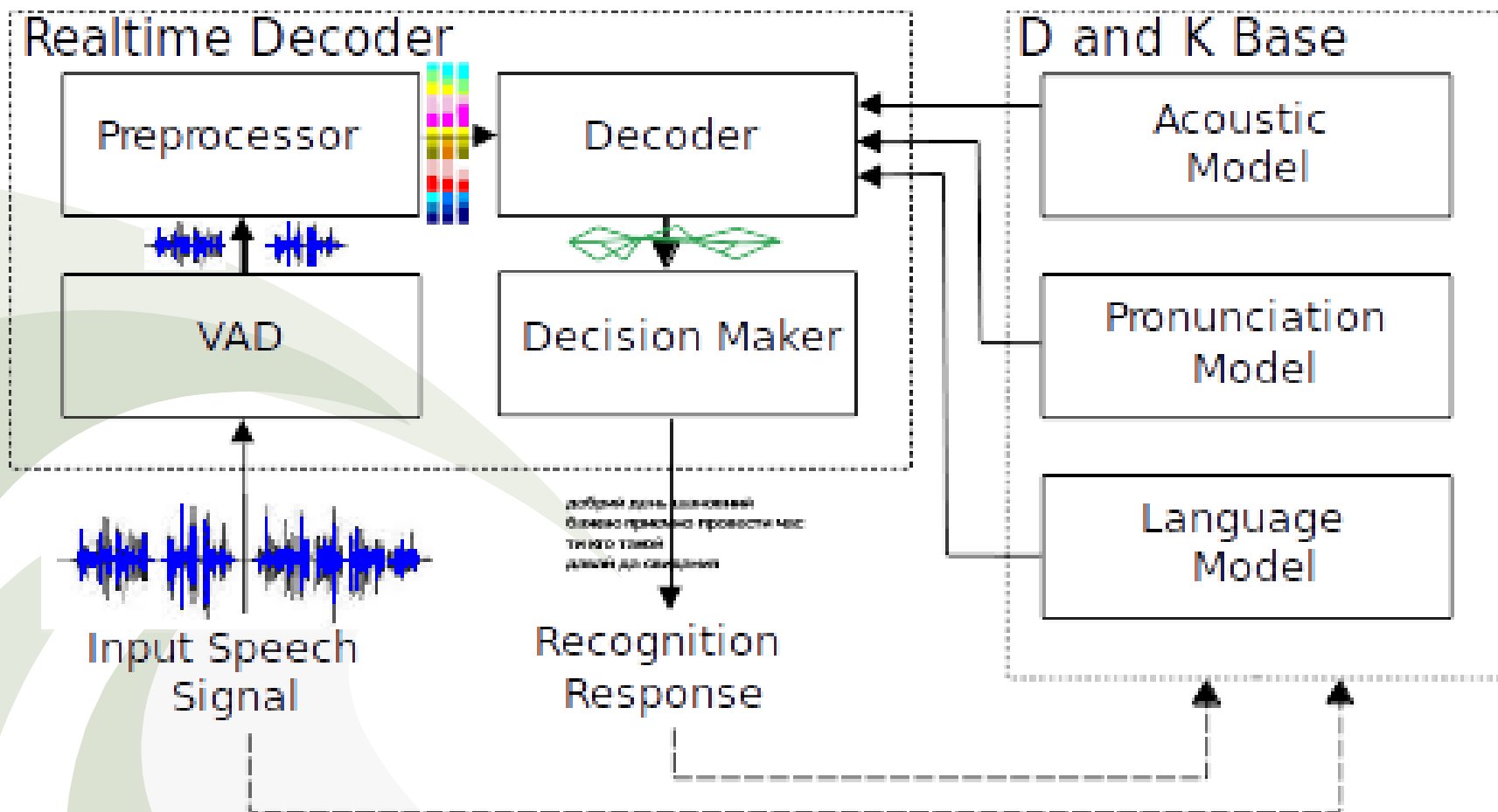
■ Синтагма, речення

■ Зміст

■ Сцена, метадані

Семантичний рівень

Загальна структура системи розпізнавання



Що говорять?



Генеративна модель як найбільш продуктивний підхід

$\mathbf{Y}_{1:T} = (y_1, y_2, \dots, y_T)$ – послідовність акустичних векторів

$\mathbf{w}_{1:L} = (w_1, w_2, \dots, w_L)$ – послідовність слів

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmax}} P(\mathbf{w} | \mathbf{Y}) \cong \underset{\mathbf{w}}{\operatorname{argmax}} p(\mathbf{Y} | \mathbf{w})P(\mathbf{w})$$

$p(\mathbf{Y} | \mathbf{w})$ – акустична складова

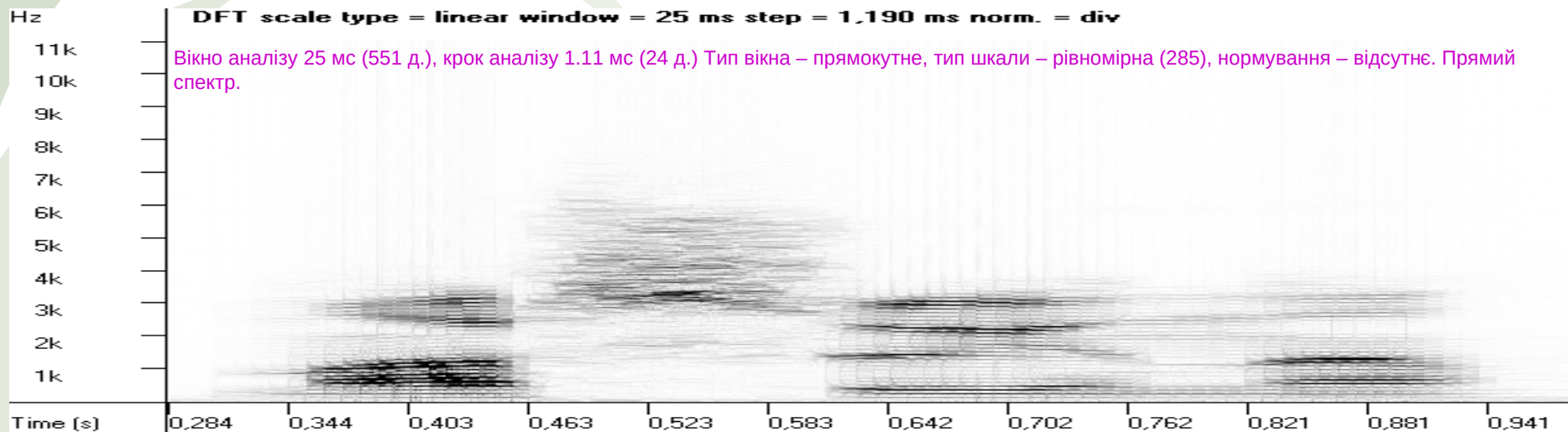
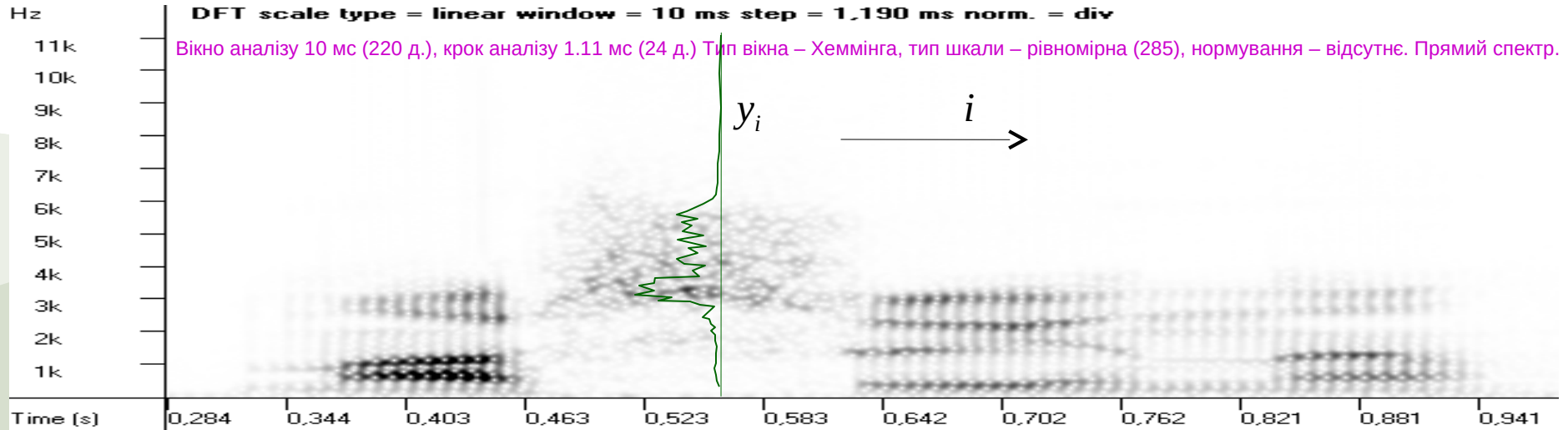
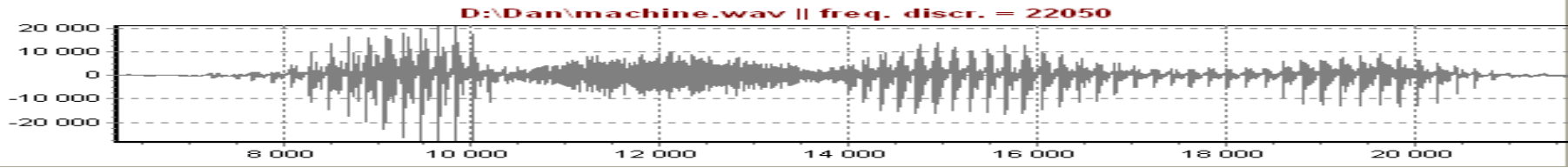
$P(\mathbf{w})$ – лінгвістична складова

Перетворення графем на фонемі

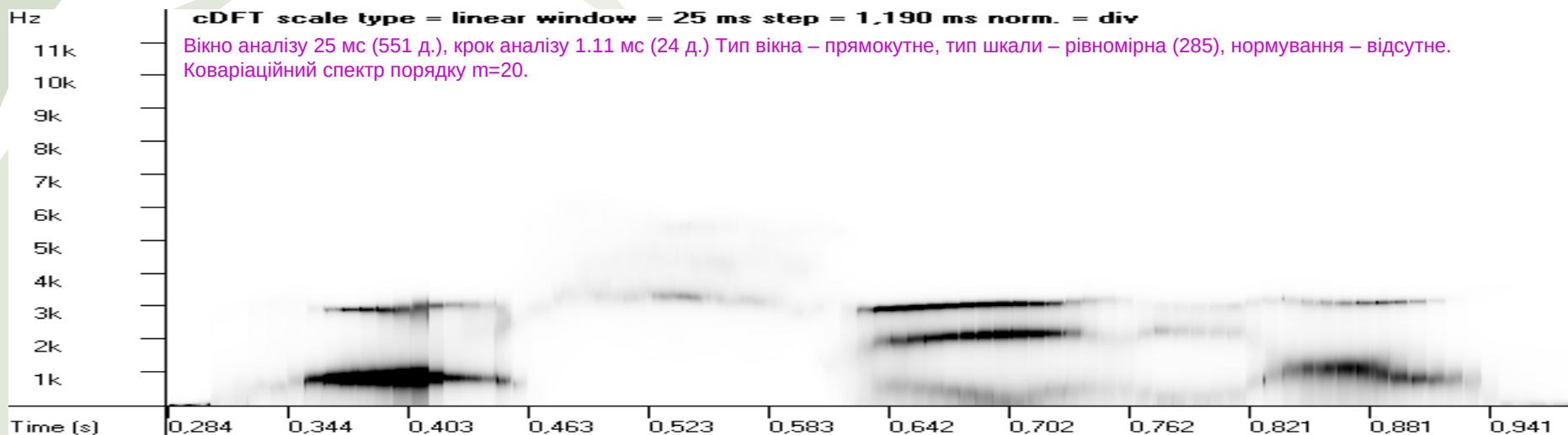
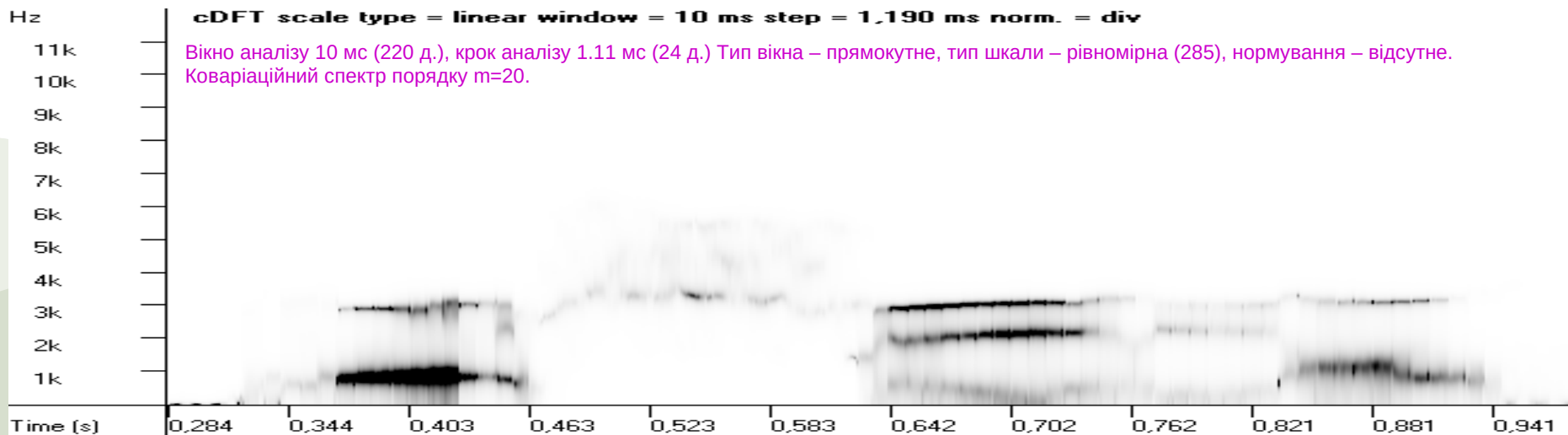
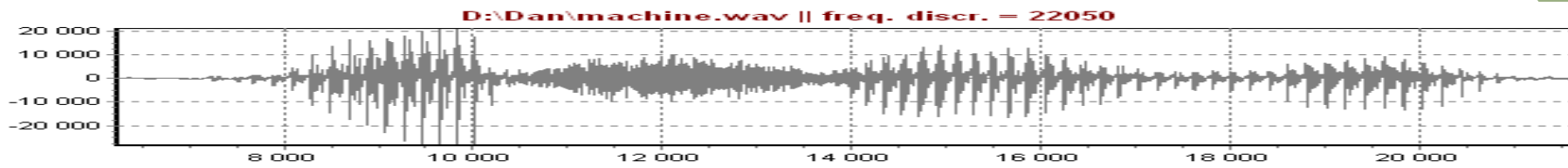
Базовий алфавіт фонем (токенів)

а а' е е' и и' і і' о о' у у'
б б' в в' г г' г' г' д д'
ж ж' з з' й к к' л л' м м' н н'
п п' р р' с с' т т' ф ф' х х' ш ш'
SIL – відсутність усної мови

Математичні моделі мовленнєвих сигналів — первинні ознаки

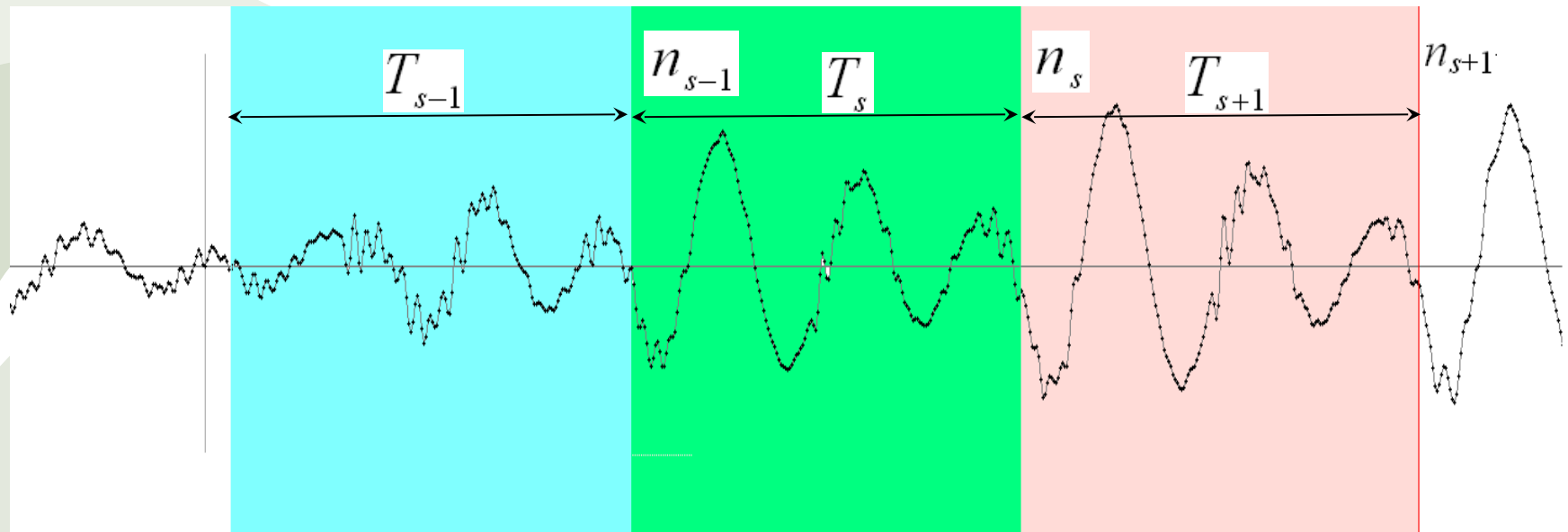


Математичні моделі мовленнєвих сигналів — первинні ознаки



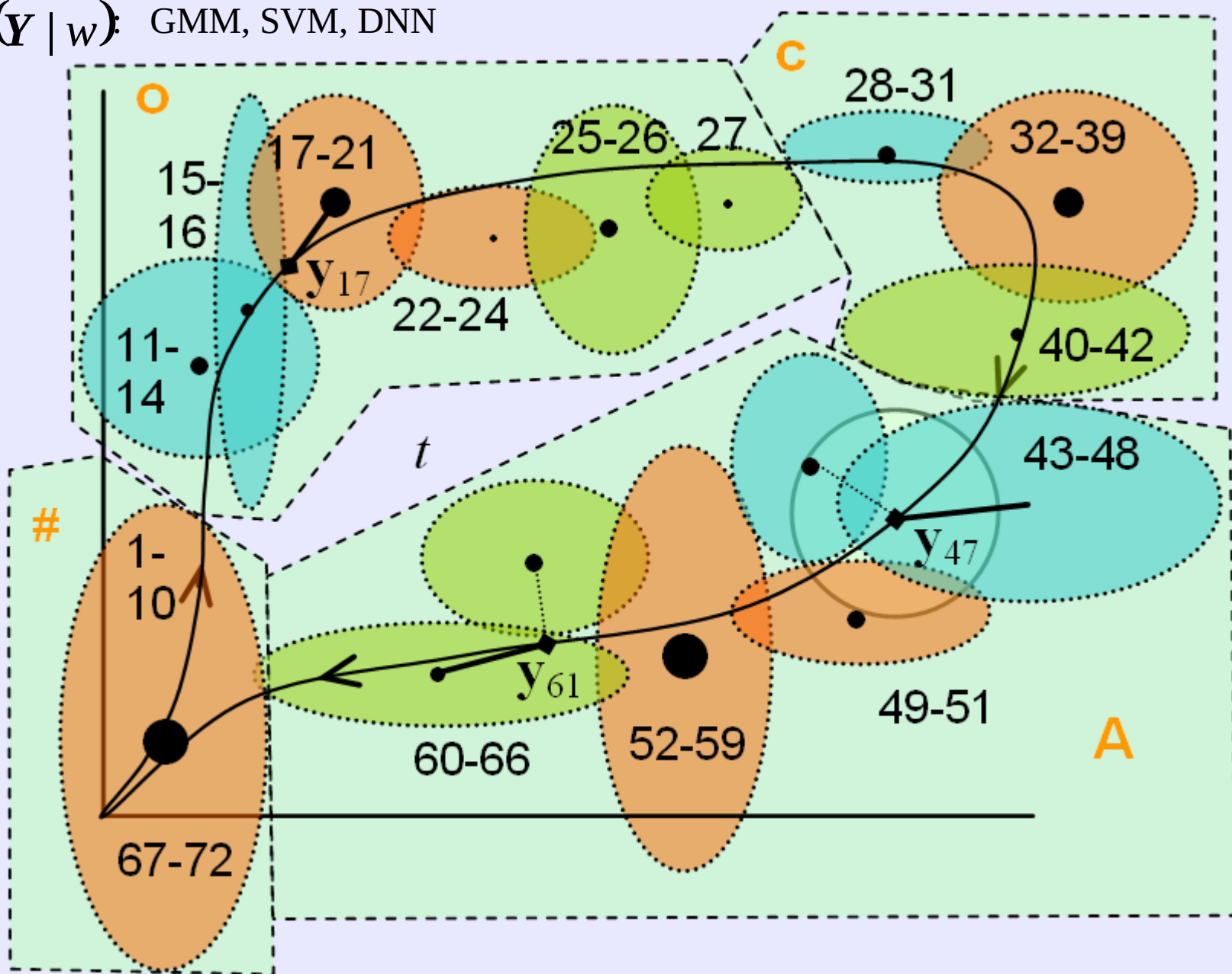
Мікрофонеми

методи автоматичного розбиття усномовного сигналу на одноквазіперіодні ділянки, квазіперіодичні та неперіодичні сегменти, розроблені методи автоматичного розпізнавання усномовного сигналу, що ґрунтуються на його попередній квазіперіодичній сегментації



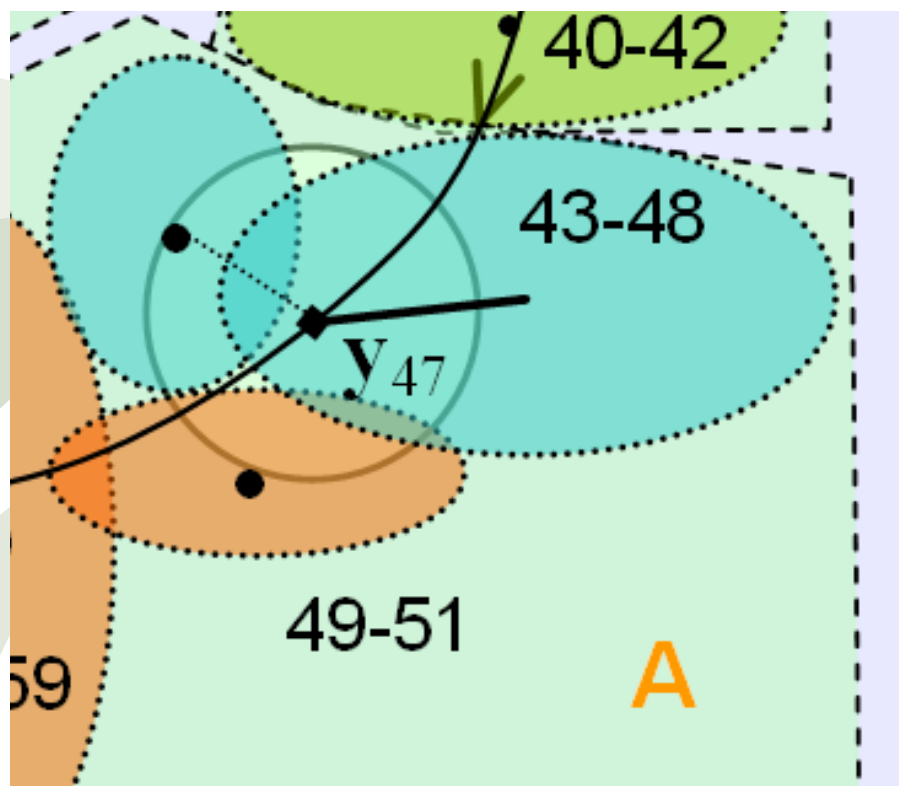
Апроксимація у просторі первинних ознак областей, що відповідають фонемам

$p(\mathbf{Y} | w)$: GMM, SVM, DNN



Апроксимація у просторі первинних ознак областей, що відповідають фонемам

$p(\mathbf{Y} | w)$ Суміші нормальних законів (гауссоїдів) – GMM



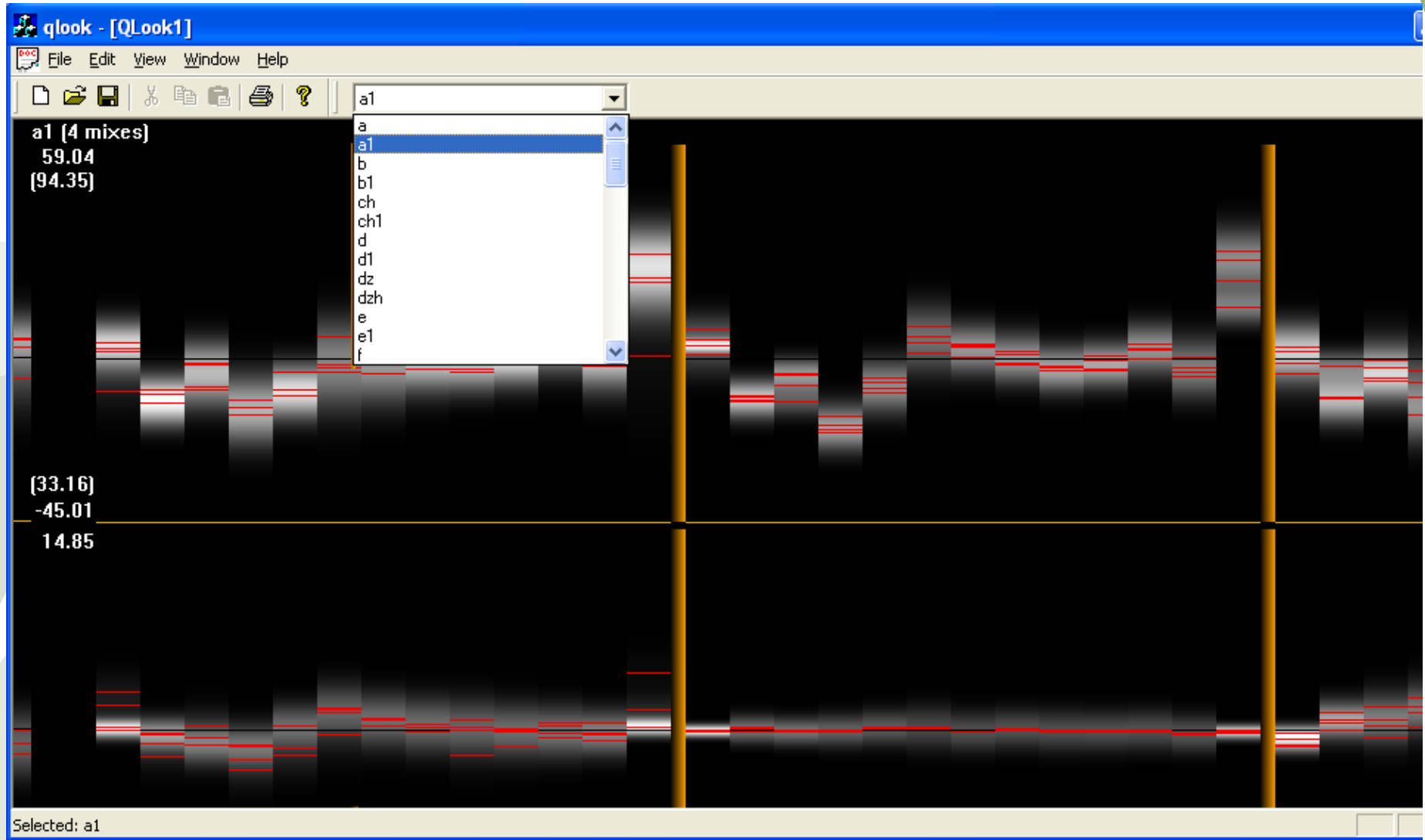
$$b_j(y) = \sum_{m=1}^M c_{jm} G(y; \mu^{(jm)}, \Sigma^{(jm)})$$

середнє

коваріація

задовольняє умовам
функції ймовірности

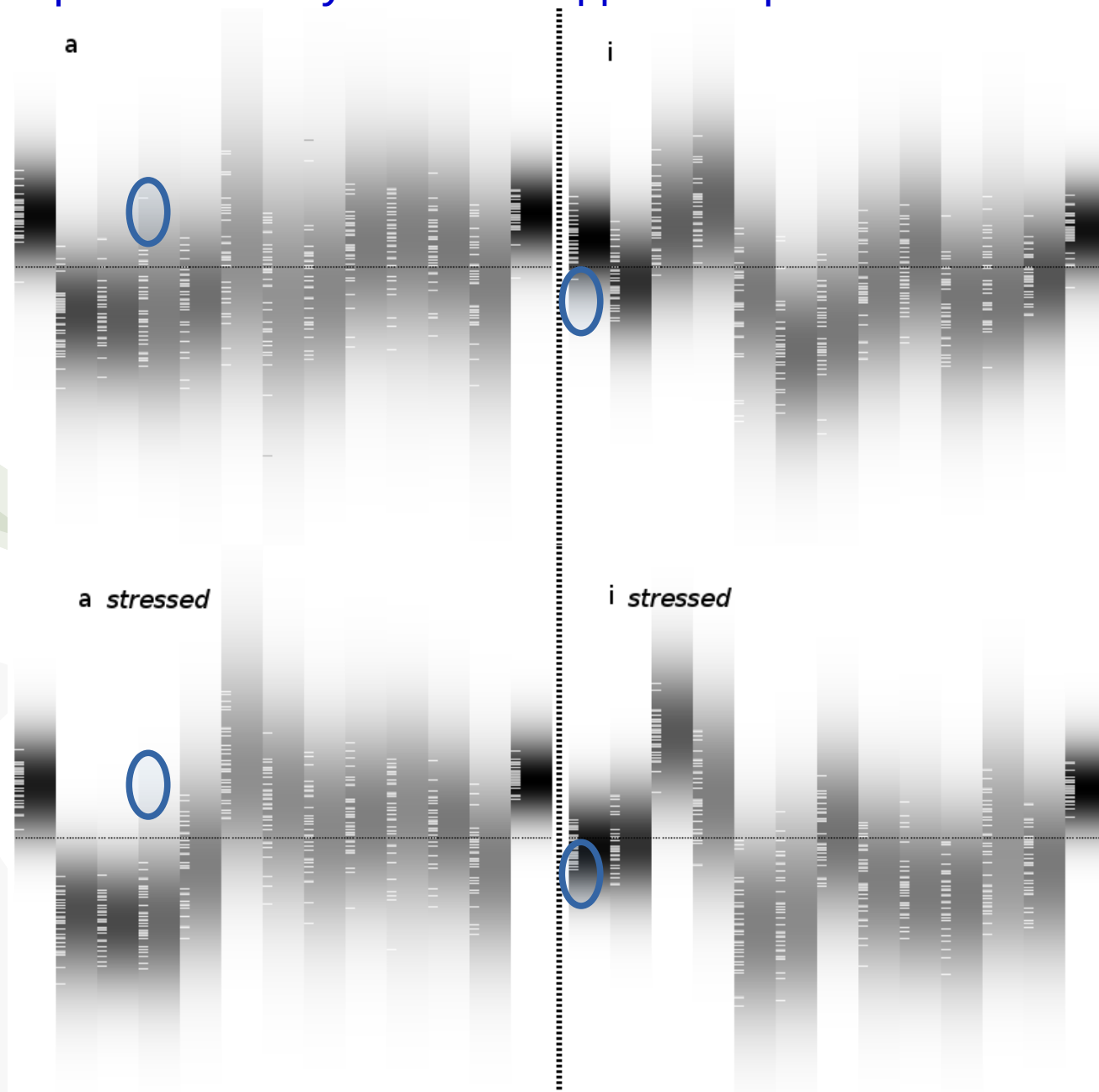
Візуалізація станів фонем — акустичний відбиток фонем



Червоні горизонтальні риси – математичні сподівання в ґауссоїді.
Висота та інтенсивність смуги відповідає дисперсії.

Візуалізація станів фонем — акустичний відбиток фонем

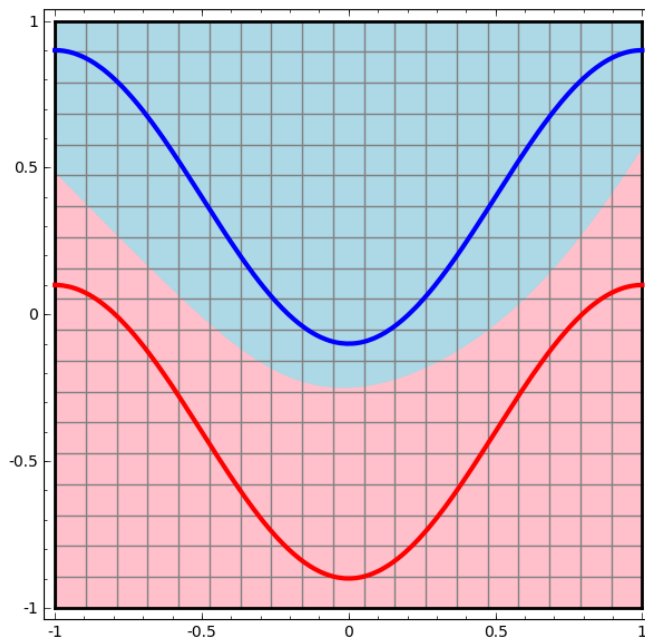
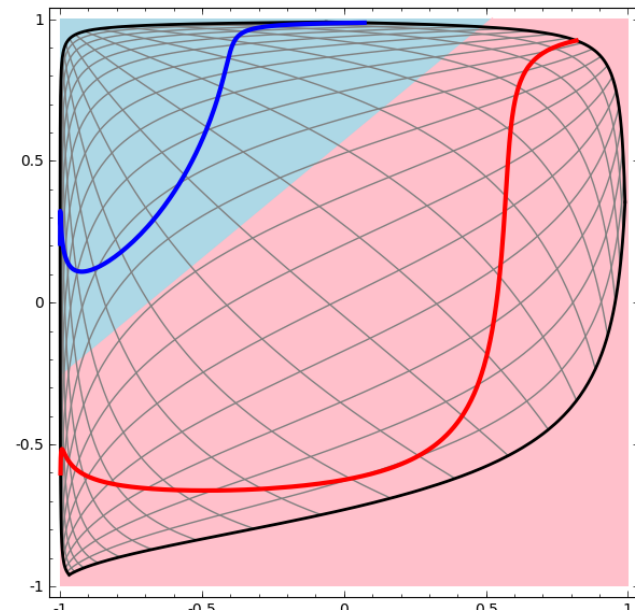
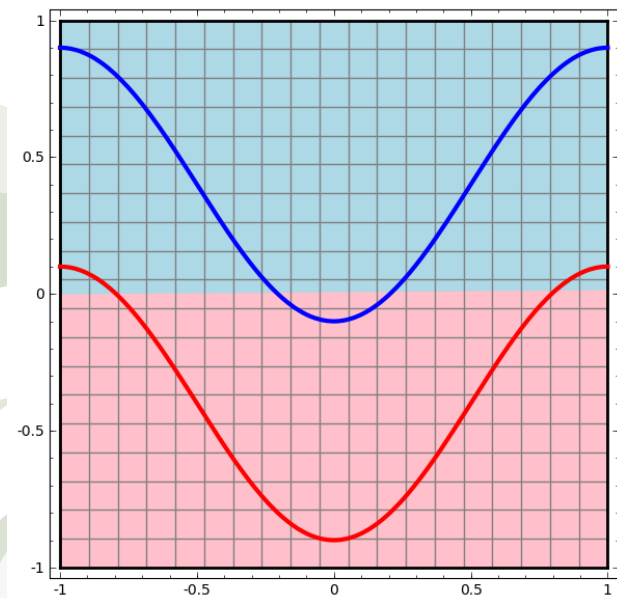
Наголос



Апроксимація у просторі первинних ознак областей, що відповідають фонемам

$$p(\mathbf{Y} | w)$$

Багатошаровий перцептрон – DNN

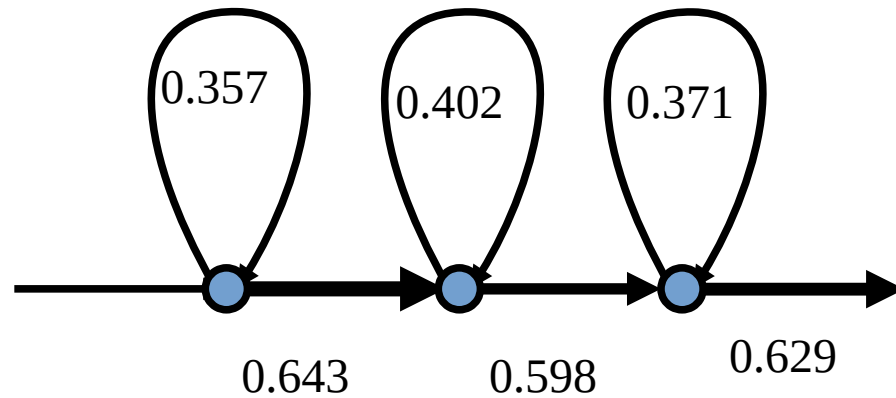


LSTM, TDNN, CTC, Transformer

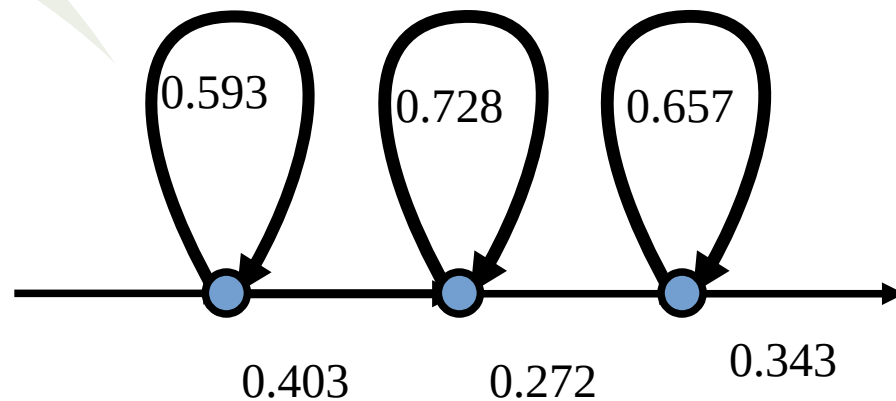
Моделювання нелінійних часових змін (базова HMM)

Жирніша петля означає триваліше перебування в стані фонемі

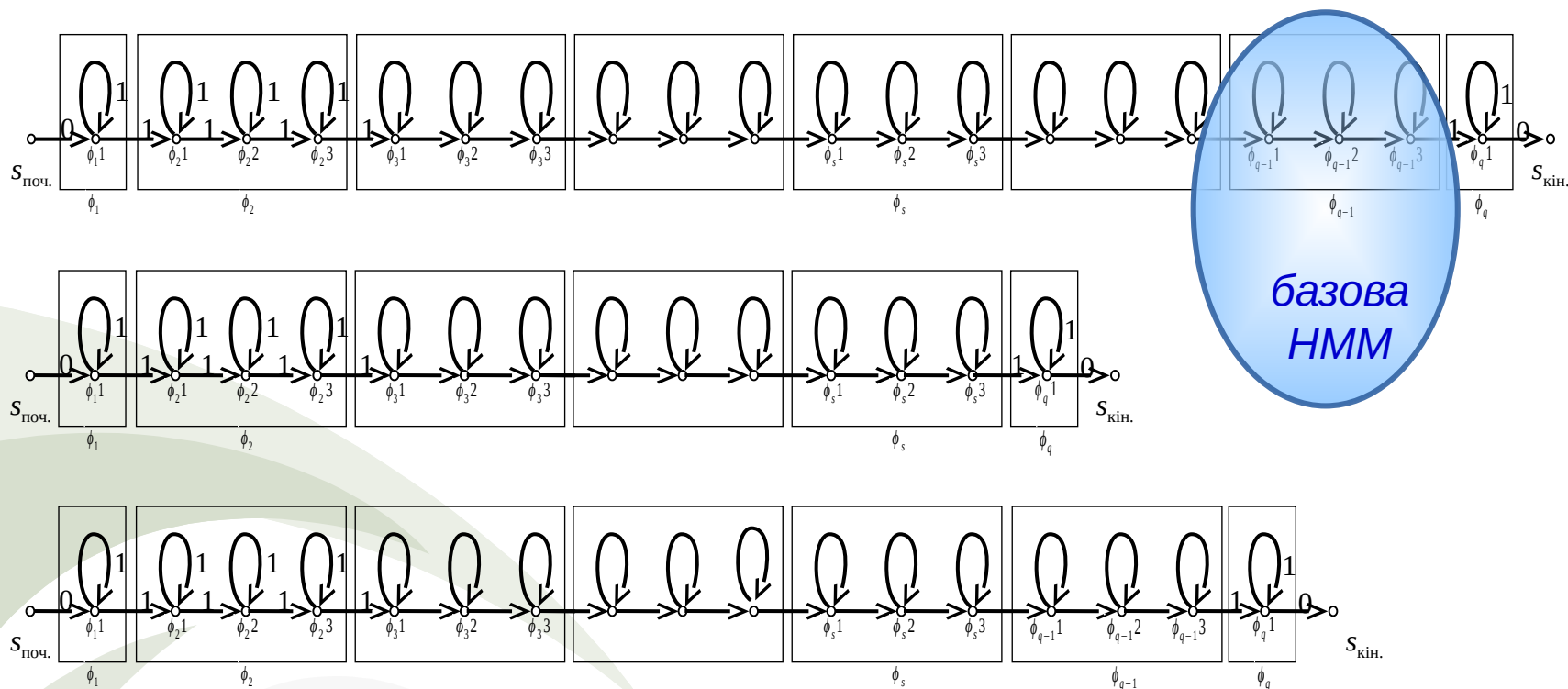
unstressed



stressed



Породжувальна граматика: компонування слів із НММ фонем



- Модельний сигнал породжується апаратом прихованих марківських моделей (*НММ — Hidden Markov Model*)
- Для кожного варіанту вимови – окрема модель слова, що компонується з акустичних моделей фонем (базових *НММ*)
- Нерегулярність в описі всієї ієрархії мовленнєвих образів
- Проблематичність моделювання явищ спонтанного мовлення

Генерування еталонного сигналу з урахуванням контексту

Нехай маємо текст, за яким генерується еталонний (модельний) сигнал:

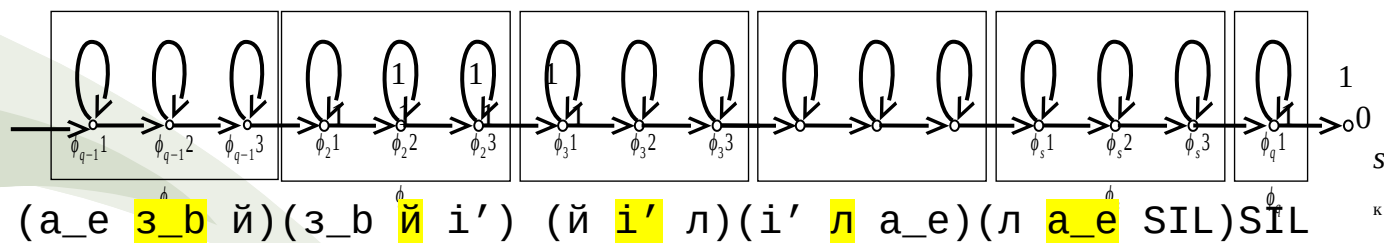
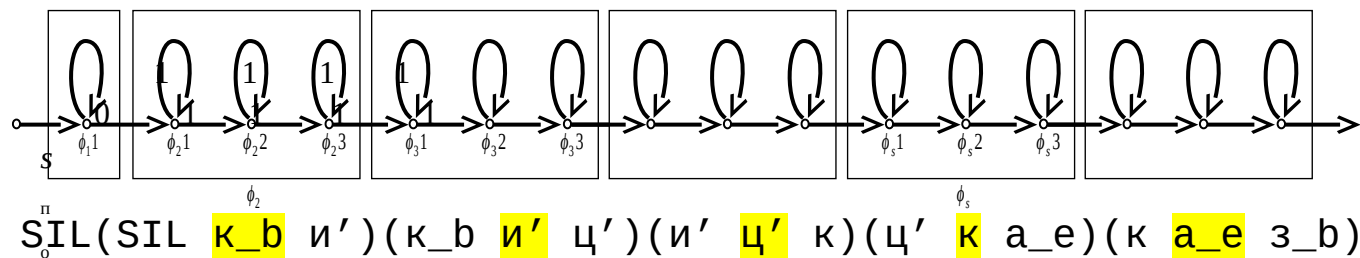
```
<s> кицька з'їла </s>
```

```
<s>    SIL  
кицька к и' ц' к а  
з'їла  з й і' л а  
</s>    SIL
```

```
SIL к_b и' ц' к а_e з_b й і' л а_e SIL
```

```
SIL (SIL к_b и') (к_b и' ц') (и' ц' к) (ц' к а_e) (к а_e з_b)  
(а_e з_b й) (з_b й і') (й і' л) (і' л а_e) (л а_e SIL) SIL
```

Генерування еталонного сигналу з урахуванням контексту



- Модельний сигнал породжується апаратом прихованих марківських моделей (*HMM — Hidden Markov Model*)
- Для кожного варіанту вимови – окрема модель слова, що komponується з акустичних моделей фонем (базових *HMM*)
- Параметри базової *HMM* оцінюються за мовленнєвою навчальною вибіркою

Лінгвістична модель: оцінювання допустимості послідовностей слів

Нехай маємо текст, за яким формується ЛМ:

$P(w)$:
n-grams

```
<s> <unk> кицька всю сметану з'їла </s>  
<s> з'їла кицька </s> ...
```

RNNLM

Контексти для ЛМ формуються у вигляді рядків, на початку яких – десятковий логарифм імовірності і послідовність слів:

\1-grams:

```
-1.0000 <unk>  
-98.9366 <s>  
-1.0000 </s>  
-0.6990 кицька  
-0.6990 всю  
-0.6990 сметану  
-0.6990 з'їла
```

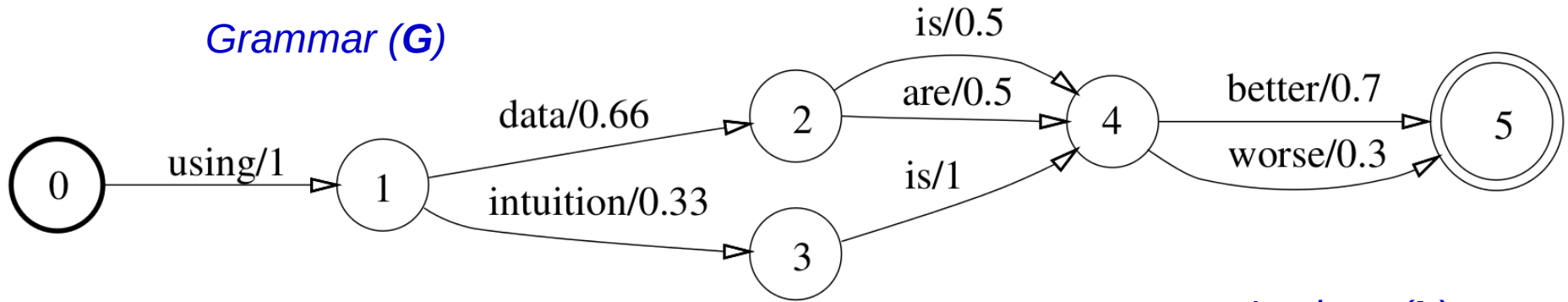
\2-grams:

```
-0.2553 <unk> кицька  
-0.2553 <s> <unk>  
-0.2553 кицька всю  
-0.2553 всю сметану  
-0.2553 сметану з'їла  
-0.5563 з'їла </s>  
-0.5563 з'їла кицька
```

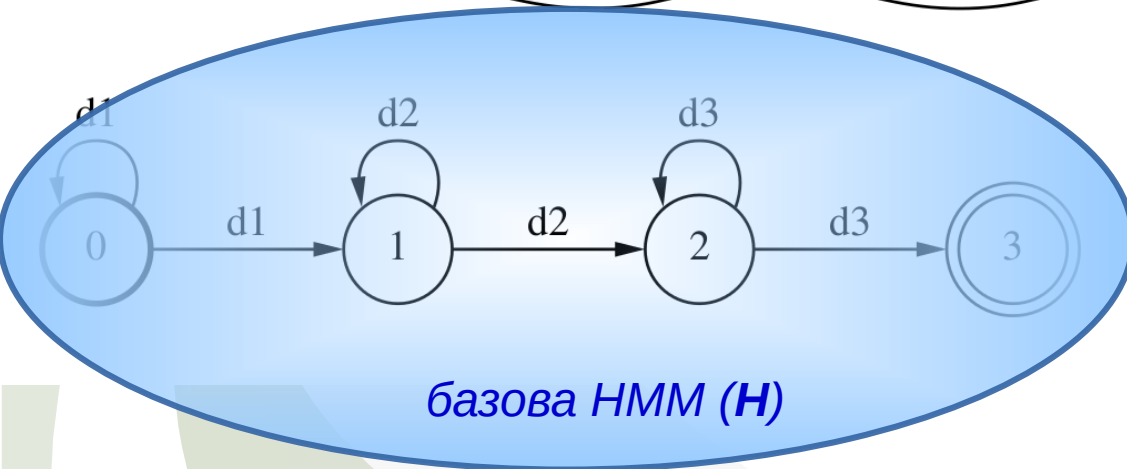
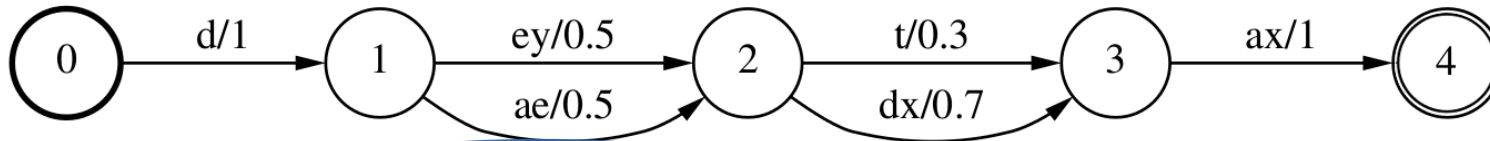
```
...  
\end\
```

Ієрархія образів: композиція трьох рівнів

Grammar (G)



Lexicon (L)



Зважені автомати / акцептори

Напрямки End-to-end та Multitask

Нехай маємо

$P(\mathbf{w})$:

n-grams

RNNLM

Формування мовленнєвого корпусу: навчальна/контрольна вибірки

Transcriber 1.5.1

File Edit Signal Segmentation Options Help

будівлю якого орендують. Досі орендна плата була символічна – одна гривня на рік, але так звана влада Криму вимагає сплачувати оренду за комерційними цінами, а це шістсот тисяч гривень.

● *муз*

Климент, архієпископ Сімферопольський і Кримський УПЦ КП

● *шт* Сьогодні культурним центром для всіх українців є саме Українська православна церква Київського патріархата. Тут і недільна школа, де діти можуть вчитися української мови і літературі, тут і культурний центр, де люди просто можуть спілкуватися".

●

Таран Лідія, ведуча

● *шт* Колядки на двадцятиградусному морозі співали в Дніпропетровську. Півсотні людей влаштували ходу центральним проспектом, щоби подарувати перехожим святковий настрій. Люди різних професій – програмісти, лікарі, землевпорядники – об'єднались через соцмережі. Всі різного віку, з різними вокальними даними, дехто приніс з собою домашню випічку, аби роздавати її містянам. Для охочих приєднатися надрукували навіть шпаргалки з текстами колядок. Святкова хода завершилася на площі героїв Майдану.

●

Гаврилюк Світлана, учасниця ходи

● *реп* Які гарні по відчуття всередині одразу, якщо вдається хоч комусь підняти трохи настрою. От люди починають посміхатися, звертати на нас увагу, хтось – може, даже приєднуватися. Це дуже класно!».

Аркушин Едуард, учасник ходи

● *реп* Треба людям нагадати, що є Бог, і він нас любить. То це дає надію. Хто в Бога не вірить, він вірить в щастя, вірить в радість, і так чи інакше – теж відчуває, бо серце є у *р* кожного, *у* душа є у кожного. І від радості і щастя сховатися неможливо».

●

Таран Лідія, ведуча

● Різдво для всіх – благодійний обід у Ідальні одного з столичних вишів влаштували для тих, хто сьогодні не може його собі дозволити. Смакували кутю, гостей розважали колядками і патріотичним вертепом, кожен ще і отримав дарунок. Детальніше – Тетяна Фруктова.

Фруктова Тетяна, кореспондент


● *реп* Святковий обід розпочинають з молитви.

● *муз*

● *реп* Серед запрошених гостей – безхатченки та пенсіонери. Киянка Людмила із подругою кажуть: на такі обіди ходять щороку, хоча

TSN_2015_01_07

Resolution



Таран Лідія, ведуча

шт Колядки на двадцятиградусному морозі співали в Дніпропетровську. Півсотні людей влаштували ходу центральним проспектом, щоби подарувати перехожим святковий настрій. Люди різних професій - ...
... надрукували навіть шпаргалки з текстами колядок. Святкова хода завершилася на площі героїв Майдану.

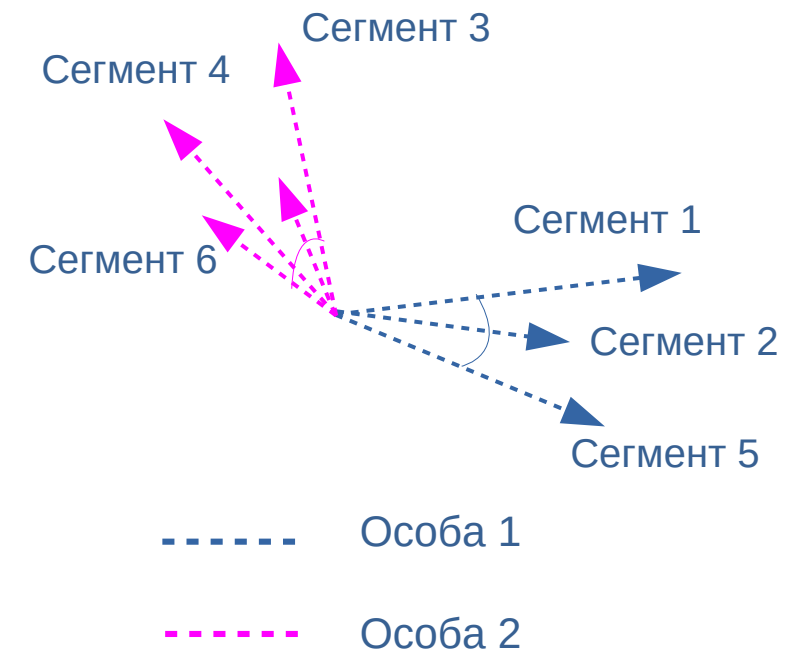
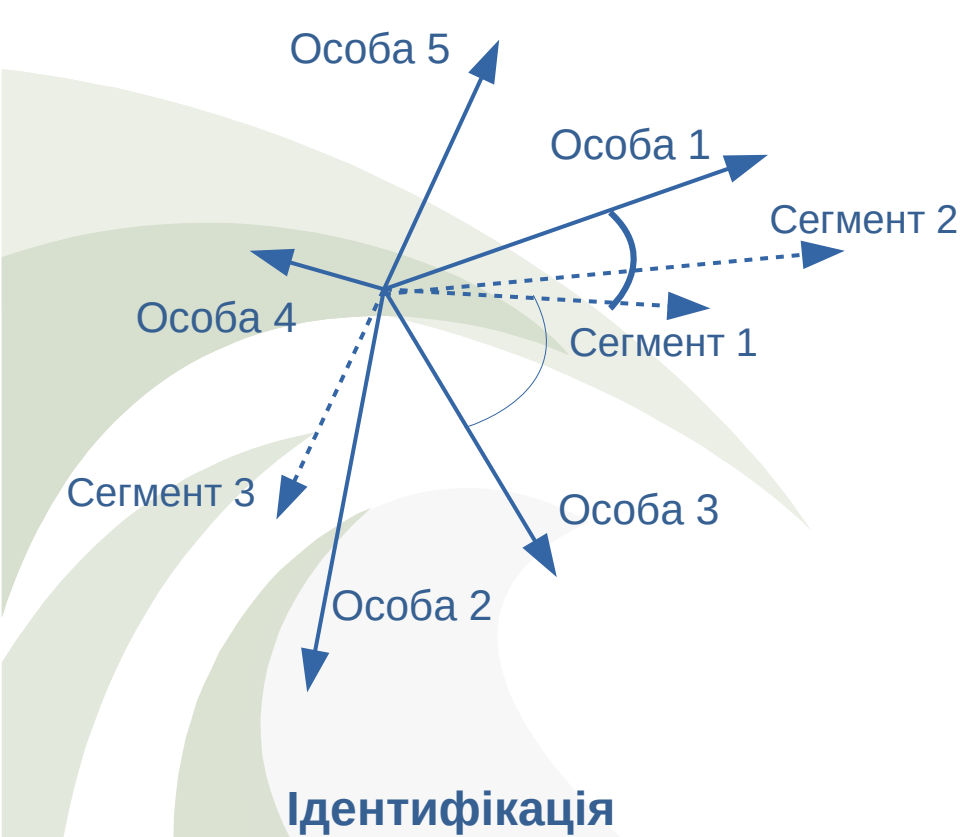
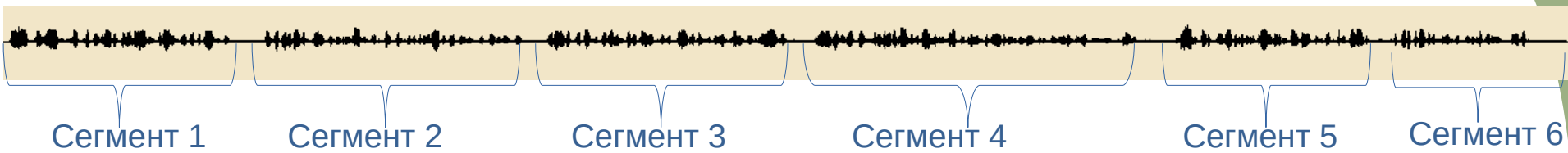
33:20 33:25 33:30 33:35 33:40 33:45

Cursor : 33:17.399

Хто говорить?



Розпізнавання дикторів методами і/х-векторів



Діаризація

Текстонезалежне розпізнавання дикторів:

- додає інформативну ознаку при перетворенні мовлення на текст;
- дає змогу оцінити кількість осіб у розмові та моменти переходу мовлення від однієї особи до іншої;
- у навчальній вибірці достатньо мати лише позначення особи, яка говорить;
- побудовані моделі не мають високої чутливості до мови;
- відкриває шлях до пошуку аудіовідеоінформації за голосом особи.

Проблеми:

- чутливість до акустичного каналу та навколишніх шумів;
- зростання помилки при збільшенні кількості осіб у розмові;
- короткі репліки, накладання голосів.

Де це працює?



Експериментальні системи: 2010-і роки

У попередніх роботах були представлені базові експериментальні системи:

- введення тексту під диктування;
- перетворення спонтанного мовлення на текст;
- моніторингу телерадіоефіру;
- усного перекладу з однієї мови на іншу;
- розпізнавання користувачів за голосом.

Експериментальна система диктування Диригент (2012)

Cybermova STT Demo - 100 000 words

File Start Finish Pause Resume

Loading Speech-to-Text data...

Done.

привіт миколо

як у тебе справи

справи йдуть непогано

ця програма раз пізнає слова і фрази

в її словнику сто тисяч слів

давайте поговоримо про погоду

зараз на дворі вітер і дощ

була сильна злива з градом

завтра буде тепло і сонце

давайте поговоримо про спорт

через два тижні почнеться євро- дві тисячі дванадцять

я купив квитки на всі матчі

але мені подобається грати в шахи

дозвольте запросити вас ресторан

краще підемо в кіно

зняли продовження фільму гарі поттер

і ми з друзями хочемо його подивитися

ми ще не говорили з вами про політику

про сумне краще й не згадувати

але восени обов'язком і підемо на вибори

і проголосуємо за тих кому довіряємо

=====

- Взаємодія з користувачем
- Покриття 95% лексики

Перетворення спонтанного мовлення на текст (2014)



5:51

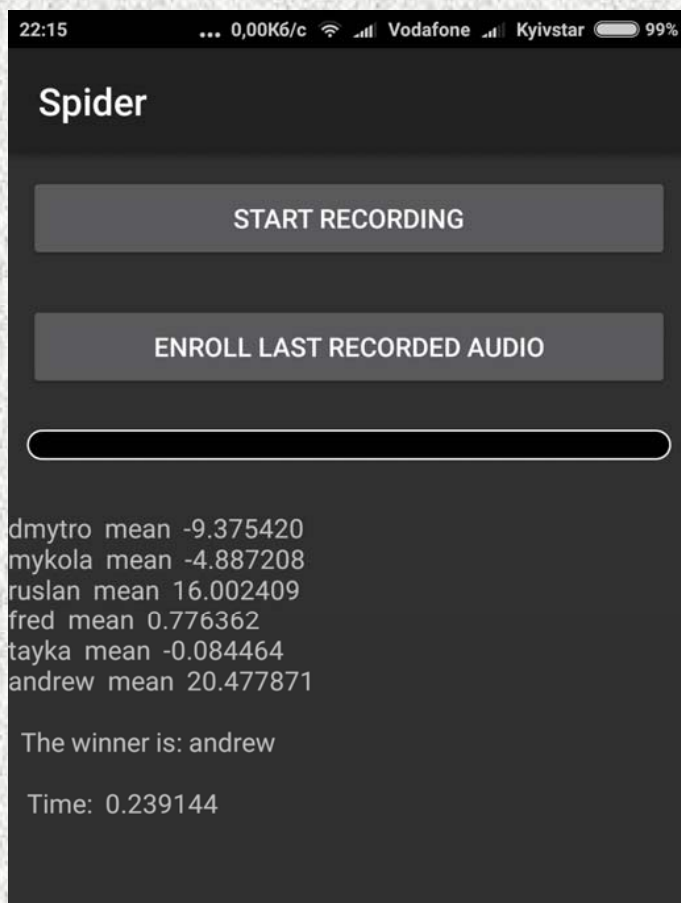


-9:26

території мусить бути притягнутий до відповідальності і не може бути цьому терміну давнини ми знаємо приклади держав, які за знищення своїх народів переслідували злочинців пожиттєво а поки останнє не відповість за те що він чинив з нашим народом так само мусить діяти україна

- Без взаємодії з диктором
- Послівна надійність 60-80%
(в залежності від якості запису)
- Веб-сервіс з можливістю редагування тексту

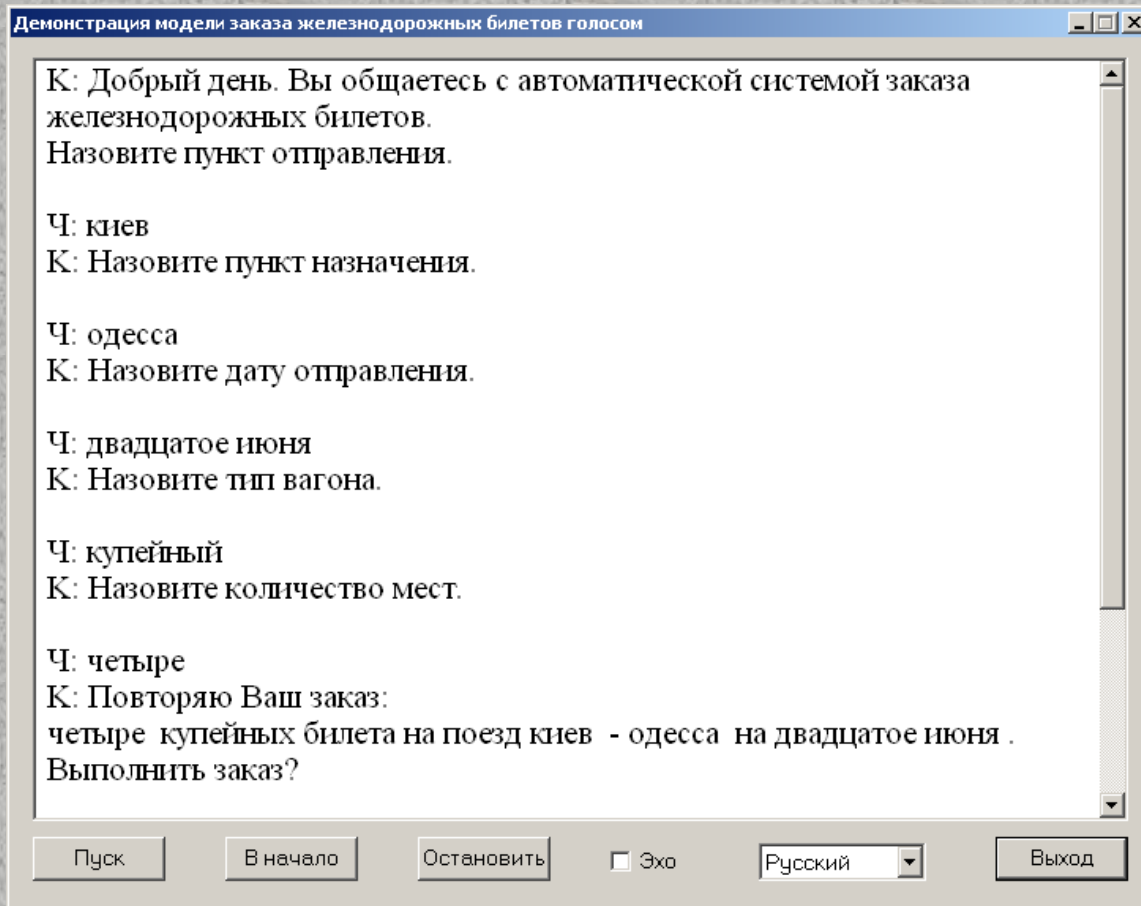
Розпізнавання користувачів за голосом (2016)



- Створення на основі моделювання індивідуальних особливостей мовця
- Визначення метаданих: хто говорить із відомих системі осіб
- Визначення моменту переходу черги говорити до іншої людини
- Реалізація як на ПК, так і на портативному пристрої

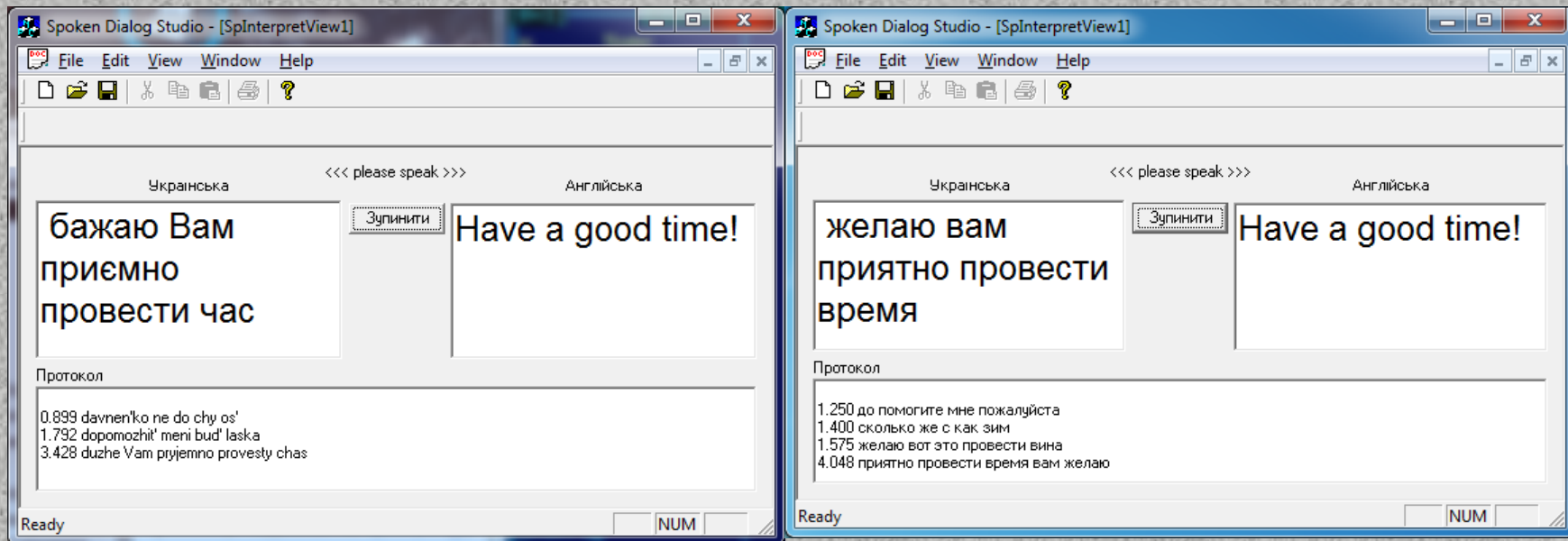
Базова система взаємодії людини і машини голосом

Підсистема усного діалогу на прикладі інтерактивного замовлення квитків



- Смыслова інтерпретація мовленнєвого сигналу
- Можливість побудови усного діалогу людини і машини

Багатомовність: усний переклад з однієї мови на іншу vOKopreter (2010)



- смисловий переклад на основі генеративної моделі смислової інтерпретації мовленнєвого сигналу
- пари з мов: українська, російська й англійська
- 14 предметних областей
-

У попередніх роботах (2010-2017 роки) були представлені базові експериментальні системи:

- введення тексту під диктування;
- перетворення спонтанного мовлення на текст;
- моніторингу телерадіоефіру;
- усного перекладу з однієї мови на іншу;
- розпізнавання користувачів за голосом.

Загальні недоліки:

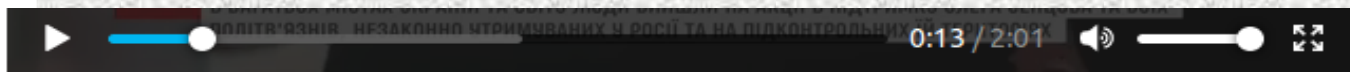
- значна чутливість до акустичних умов, відмінних від умов у навчальній вибірці;
- відносно невеликий словник (OOV>3-5%);
- при перетворенні мовлення на текст індивідуальні особливості не моделювалися;
- відсутність реальної багатомовності.

Сучасні системи:

- перетворюють на зрозумілий людині текст практично будь-яку звукову медійну інформацію;
- мають кращу надійність - в середньому 75-95% - і меншу чутливість до акустичних умов;
- прогнозують розділові знаки;
- оцінюють ступінь довіри до окремих розпізнаних слів;
- відображають ознаки індивідуальності;
- моделюють реальну багатомовність;
- активно експлуатуються!

Дослідна експлуатація

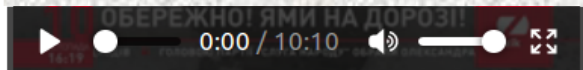
Результат перетворення на текст звукової медійної інформації з новинного каналу ТСН



П'ЯТДЕСЯТИЙ ДЕНЬ ГОЛОДУВАННЯ ОЛЕГА СЕНЦОВА УКРАЇНСЬКОГО РЕЖИСЕРА ПОКИ НЕ ГОДУВАННЯ
ПРИМУСОВО ПРО ЦЕ ЗАЯВИВ АДВОКАТ БРАНЦЯ КРЕМЛЯ ДМИТРО ДІНЗЕ ОДНАК ВІН НЕ ВИКЛЮЧИВ
ТАКОГО КРОКУ РОСІЙСЬКІ ТЮРЕМНИКИ МОЖУТЬ ВДАТИСЯ ЯКЩО СТАН СЕНЦОВА ПОГІРШИТЬСЯ
ПЕРЕБУВАЄ У МЕДИЧНІЙ ЧАСТИНІ В'ЯЗНИЦЯ БЛЯ ПОЛЯРНОГО КОЛА І НЕ МОЖЕ ХОДИТИ А ТИМ
ПІДТРИМКУ УВ'ЯЗНЕНОГО РЕЖИСЕРОМ БАГАТЬОХ МІСТАХ УКРАЇНИ ТА СВІТУ ЛЮДИ ВИЙШЛИ НА
ЙОГО ПІДТРИМКУ ГАННА БОКА ДАЛІ І ЙОГО МИ З ТОБОЮ МИРЕ МИ С ТОБОЮ ЛЬВІВ ДНІПРО ОДЕСА І
УКРАЇНСЬКИХ МІСТ ПРИЙМАЄМО НАШИХ ОЛЕГА АВСТРІЯ ФРАНЦІЯ І ЩЕ ЗО ДВАДЦЯТЬ КРАЇН СВІТУ
НІДЕРЛАНДИ НА ЗВ'ЯЗКИ МИ ЗНАХОДИМОСЯ ПОЗАДУ ТРИБУНАЛУ УСЮДИ УКРАЇНЦІ ПО ВИХОДИ
МАЙДАНИ ПІДТРИМАТИ ОЛЕГА СЕНЦОВА ТА ВСІХ В'ЯЗНІВ КРЕМЛЯ ВУЗЬКА СІ ВОСЬМИ-ДЕСЯТИ ЛЮДЕЙ
ЗНАХОДЯТЬСЯ В ЗАРУЧНИКАХ В РОСІЙСЬКИХ В'ЯЗНИЦЯХ І МИ МАЄМО ЗБИРАТИСЯ І МИ МАЄМО ПІ
ГОВОРИТИ І УКРАЇНА ГОТОВА ПЕРЕДАТИ МОСКВІ ДВАДЦЯТЬОХ ТРЬОХ ЗАСУДЖЕНИХ РОСІЯН В ОБМІН
СТІЛЬКОХ УКРАЇНЦІВ ЗАГ'РАТОВАНИЙ КРЕМЛЕМ ІЗ ПРОХАННЯМ ПОМИЛУВАТИ сенцова ВЖЕ ЗВЕР
ПУТІНА І ГЕНСЕК РАДИ ЄВРОПИ ТУРБІОРН ЯГЛАНД ТА У КРЕМЛІ ЛИШЕ ВІДПОВІЛИ ПОМИЛУВАТИ
НА ПІДСТАВІ ЦЬОГО ПРОХАННЯ НЕ МОЖУТЬ ПРОСТІЙ Ж УКРАЇНЦІ ПЕВНІ ЯКЩО МІЖНАРОДНА ПІ
БУДЕ НЕ РАЗОВОЮ НАШИХ СПІВГРОМАДЯН ІЗ РОСІЙСЬКИХ КАТИВЕНЬ ВДАСТЬСЯ ВРЯТУВАТИ ФРІ
СТАЄ БЕЗОСТРОКОВЕ УКРАЇНЦІ В КИЄВІ ТА ПО ВСЬОМУ СВІТУ ВНИМАЮТЬ І ПОСОЛСТВО ФРАНЦІЇ ТА

Дослідна експлуатація

Синім кольором позначено слова з нижчим ступенем довіри



0

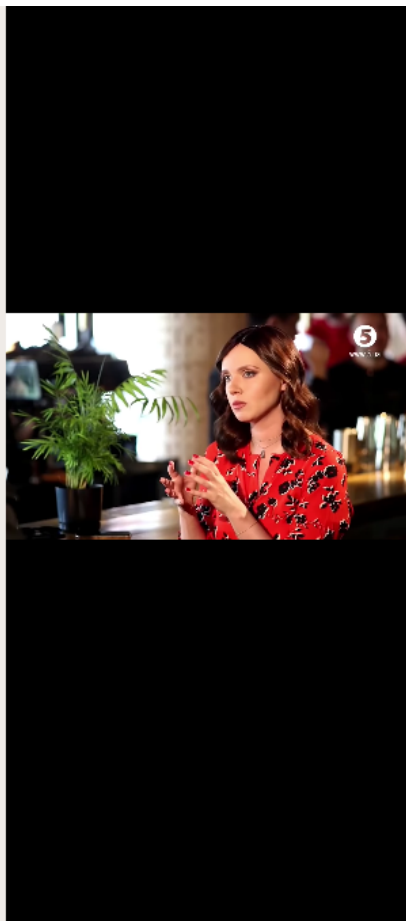
НАВІТЬ У ВОДУ НАВІТЬ ДОТЕПНІШИЙ ЗНАХОДЯТЬ ВИХІД ІЗ СИТУАЦІЇ А ВОДІЇ ВЖЕ НАВІТЬ НЕ СПОДІВАЮТЬСЯ ПОБАЧИТИ РІВНІ ДОРОГИ В УКРАЇНІ ПОХІД У РОЛІ ЦЕ Є ДОРОГИ СКАЖІТЬ МЕНІ БУДЬ ЛАСКА ЗВІЛЬНИ ДОРОГИ СЕЗОНУ МОЖНА виражатися не можна хреновее ЦЬОГОРІЧНЕ ЛІТО ПЕРЕВЕРНУЛО ВСЕ З НІГ НА ГОЛОВУ ПОД РАЗУ В ДВІСТІ КІЛОМЕТРІВ ЗАПОРІЖЖЯ МАРІУПОЛЬ ЗРОБИЛИ ЗА ЛІЧЕНІ МІСЯЦІ ДЕКІЛЬКА РОКІВ ЇЇ РОБИТИ А ПОТІМ РАЗ ТАМ ОБІЦЯНКА ПОЛІТИЧНА ОБІЦЯНКИ І РАЗ ВОНА ТАМ ЗА ДВА МІСЯЦІ ЗА ЗАВЕРШЕНО З ПОЧЕСТЯМИ ПЕРШИМ ПО НІЙ ПРОЇХАВ ПРЕЗИДЕНТА ЯКЩО ВСІ ДОРОГИ ВИРІШИТИ ЧОТИРИ РОКИ ЧЕРЕЗ ДВА РОКИ ОТЖЕ НАЙВАЖЛИВІШІ ТРАС НАШОЇ КРАЇНИ ДОБРЕ ЯКЩО ТАК КАЖУТЬ ЕКСПЕРТИ І ПРИГАДУЮТЬ ОБІЦЯНКИ ЯКНАЙШВИДШЕ ВІДРЕМОНТУВАТИ ДОРОГИ ВЖЕ ТРАДИЦІЙНО ЧУЄМО ВІД УСІХ ВИСОКОПОСАДОВЦІВ ПРОГРАМА ДОРОЖНЬОГО БУДІВНИЦТВА ЦЕ СТРАТЕГІЧНА ЛІНІЯ УКРАЇНСЬКОЇ ВЛАДИ ТЕПЕР МОЖЕ ПРОВЕСТИ ХОРОШИЙ порядок на дорогах ЦЬОГО РОКУ НА БУДІВНИЦТВО ДОРОГИ ПЛАНУЮТЬ НОВИЙ РЕКОРДНИЙ БЮДЖЕТ ПОНАД СІМДЕСЯТ МІЛЬЯРДІВ ГРИВЕНЬ ТА ЧИ ВИСТАЧИТЬ НА РІВНЕНЩИНІ асфальт В УКРАЇНІ БЛИЗЬКО СТА СІМДЕСЯТИ ТИСЯЧ КІЛОМЕТРІВ ДОРІГ ТУТ ДВАДЦЯТЬ ТРИ ТИСЯЧІ КІЛОМЕТРІВ ЦЕ ДОРОГА МІСЦЕВОГО ЗНАЧЕННЯ І СОРОК СІМ ТИСЯЧ КІЛОМЕТРІВ МІЖНАРОДНІ ДОРОГИ ТАК ЗВАНІ ТРАНЗИТНІ КОРИДОРІ ДЛЯ ТОГО ЩОБ ВІДРЕМОНТУВАТИ ПРОСТО ЗБЕРЕГТИ ЦІ ДВАДЦЯТЬ З ЛИШКОМ ТИСЯЧ НАЦІОНАЛЬНИХ ДОРІГ ВСЕСВІТНІЙ БАНК У СІМНАДЦЯТОМУ ЗДАЮТЬ ЗДАЄТЬСЯ РОЦІ РАХУВАВ І ВОНИ НАРАХУВАЛИ ЦЕ ПІСЛЯ ЧОГО МІСЦЕВІ СЛУЖБИ СКАЖУТЬ ЩО БУДЬОСЬ НЕ ПІДІЙДЕ ТАК



Автоматичне двомовне оцифрування телеканалів відбувається сегментами по 30 хвилин

РОЗПОВСЮДЖУЮТЬСЯ КОНВЕНЦІЯ НА ЇХ ВІЙСЬКОВУ ПРИСУТНІСТЬ І ВІЙСЬКОВУ АКТИВНІСТЬ В ЧОРНОМУ Й АЗОВСЬКОМУ МОРЯХ ТОМУ ВОНИ ВВАЖАЮТЬ ЩО НЕ МАЄ ЮРИСДИКЦІЮ ТРИБУНАЛУ РОЗГЛЯДАТИ ЦЮ З ЦЮ СПРАВУ НУ ОТ ЗАРАЗ ТРИБУНАЛ МАЄ ВИЗНАЧИТИСЬ ЯКЩО МИ ЯКЩО ВОНИ ПРИЙМУТЬ ТРИБУНАЛ ПРИЙМЕ РІШЕННЯ ЩО СПІР ПОДІЛЯЄТЬСЯ НА ДВІ ФАЗИ ТАК ЗВАНУЮ БІФУРКАЦІЯ ПРОЦЕСУ ТО БУДЕ СПОЧАТКУ РОЗГЛЯД ПИТАННЯ ПО ЮРИСДИКЦІЇ ЧИЙ ЮРИСДИКЦІЄЮ ТРИБУНАЛУ РОЗГЛЯДАТИ ЦЕЙ СПІР А ПОТІМ ВЖЕ БУДУТЬ РОЗГЛЯДАТИСЯ ПИТАННЯ ЩО СТОСУЮТЬСЯ ТОГО В ЧОМУ САМЕ ПОЛЯГАЮТЬ ПОРУШЕННЯ І ЯКА КОМПЕНСАЦІЯ ЗА ЦІ ПОРУШЕННЯМ МОЖЛИВО НУ ОТ МИ ЗАРАЗ НА ЕТАПІ КОЛИ ЦЕ МАЄ БУТИ ВИЗНАЧЕНО І КОЛИ МИ ДІЗНАЄМОСЬ ЯКІЙ НАС БУДЕ ПОДАЛЬШИЙ РОЗГЛЯД СПРАВИ хорошо хотел его спросить по поводу евроинтеграции а вы сами сказали что в начале нашего разговора что много занимались вопросам связанными с подготовкой украины к подписанию договора о европейской интеграции зоны свободной всеобъемлющей какая живая чужой зоне свободной торговли и я прекрасно помню очень тринадцатого года когда многим казалось что все вопрос уже решен потому ты была такая реакция в ноябре когда вопрос неожиданно оказался совсем не решен на мои развернулся на сто восемьдесят градусов тогдашний президент и тогдашний премьер-министр А ОТ но главное что в то время было огромное количество СТРАШИЛОК что все вот политическая интеграция будет означать то что украины попадет в политическую кабалу и евросоюзу а зоны свободной торговли всеобъемлющее приведет к тому что украинская экономика рухнет и развалится и в общем здесь будет какой-то сплошное пепелища но как мы понимаем что ничего этого не случилось несмотря на войну да но все-таки

Діаризація: розмова розбивається на сегменти, що відповідають різним особам, яким ставиться у відповідність номер



3: нашому ефірі

2: ні кримські татари йдуть такі знають як процесуально наступом бо нам треба об'єднуватися не лише між собою але вибачте це можливо пафосну нам треба об'єднувати її розрізнений колони українці в якій ждуть в одному напрямку демократичному європейському отже цьому складі парламенту я дуже сподіваюся нас буде не два депутати кримських татар а можливо буде п'ять

3: наскільки мені відомо і глядачам зараз пояснимо що а кримські татари є в інших партіях друзі а в списках партії європейська солідарність петра порошенка а будуть представлені мустафа джемільєв заступник глави меджлису ахтем чийгоз списках партії українська стратегія яку очолює нагадаю вам а прем'єр-міністр володимир гройсман перший заступник міністра інформаційної політики буде а імена джапарова партія голос нагадаю святослав вакарчук а очілює цю партію буде два представники кримських татар на третій позиції у списках сила і честь ігоря смешка власне рефат чубаров знаєте я вам відверто скажу для мене дивно що якщо ми маємо а там об'єднану кримськотатарську опозиції а у вас вона об'єднана і би доволі солідарні всі разом а що ось таким чином заповнені яєчки складів партії які принципи подекуди мають різний характер векторний щодо ігоря смешка у чому саме вибір ваш припав на ігоря з мішками на солідарність порошенка з ним вас пов'язують досить непогані стосунки там гройсман вакарчук новоспечений

2: коли спарты но хоть и не все так ну скажем такую идею стороною як правило багатофакторний такі перемовини і тут особливо дуже важливим є такою розуміння того хто і навіщо об'єднуються на завод декілька ключових с ним

3: в особняке

2: які нас об'єднують здається я дуже хотів щоб ніхто не подумав що мовляв крім татар таких знайдете хитруючи чому не розкласти яйця по всі рекорди на таксі

3: так і подумала я думаю у вас на вас питаю

2: європейський союз нато де оккупация криму і національно-

00:01:35 = 95.74s / 00:10:00

Open

Connect

Continuous

Synchronize



1.0x

Перспективи:

- передбачення пунктуації з урахуванням просодики (інтонації та ритму) та багатомовності;
- реалізація пошуку особи за голосом у масиві фонограм;
- сегментація на епізоди;
- визначення тональності та токсичності мовлення з урахуванням просодики;
- пошук шумів певних типів (сміх, оплески тощо);
- реконструкція тексту;
- мультимодальність.

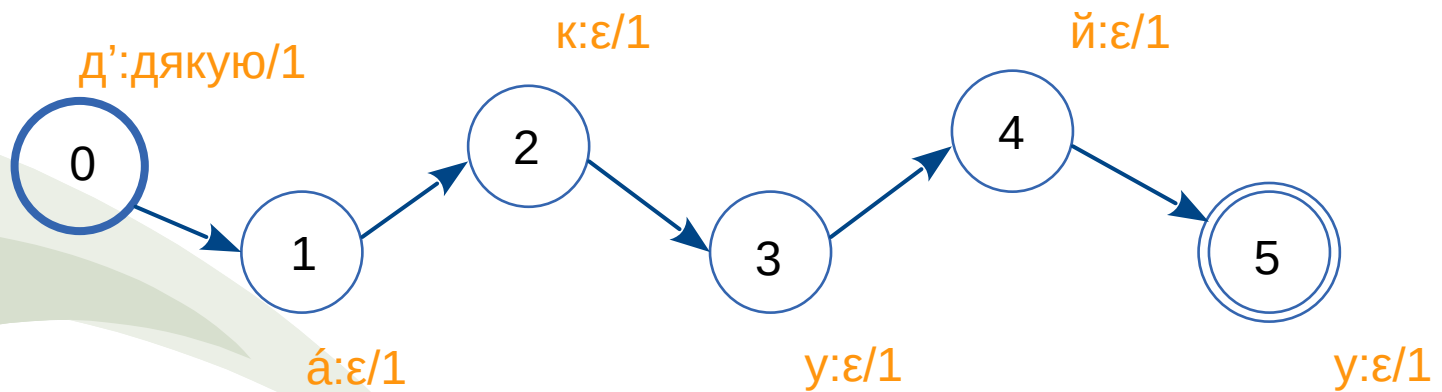
Періодичні наукові заходи

- IEEE: International Conference on Acoustics, Speech, and Signal Processing (ICASSP) і менші конференції та семінари;
- InterSpeech (рік у Європі, рік у світі), International Speech Communication Association (ISCA);
- Odyssey: The Speaker and Language Recognition Workshop
- Conference on Text, Speech and Dialogue (TSD), Czechia
- Під егідою Int. Association for Pattern Recognition (IAPR);
- ...

Приклади комерціалізації

- Диктування:
 - Nuance (Dragon Dictate) — Apple, Google — Android, Microsoft, Baidu
- Транскрибування фонограм (www.pcmag.com/picks/the-best-transcription-services):
 - Speechmatics, rev.com, otter.ai, verbit.ai, sonix.ai, ...
- Телерадіоефір (транскрибування, субтитрування):
 - Autonomy — Hewlett Packard — Microfocus: система IDOL (Intelligent Data Operating Layer), HappyScribe, Digital Nirvana
- Автоматизація кол-центрів:
 - Omilia, Nuance
- Анотування мітингів, продуктивність команд:
 - Sembly.ai, Lucid Meetings
- Системи з відкритим кодом:
 - OpenAI, HuggingFace, ряд комерційних компаній, наукова спільнота (Kaldi, EspNet)
- Різне:
 - Facebook, Amazon, NVidia
- Загальний обсяг ринку оцінюється десятками млрд. доларів

Дякую за увагу!



info@cybermova.com

Літні школи-семінари, 2004-2011 роки



с. Жукин, річка Десна (2004)

“Над нами завжди хмарне небо” :)