

NEUROSOFT® 

NeuroVoice

Synteza i analiza mowy

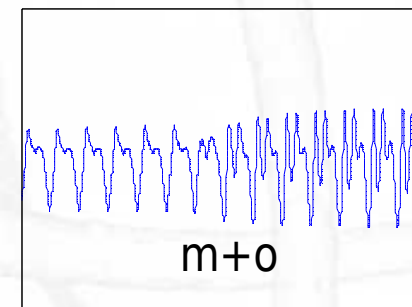
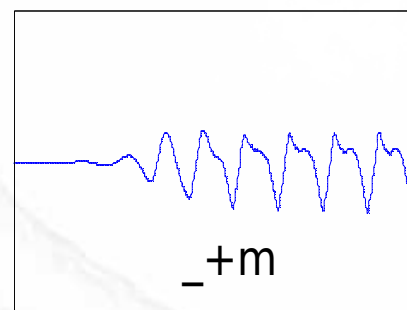
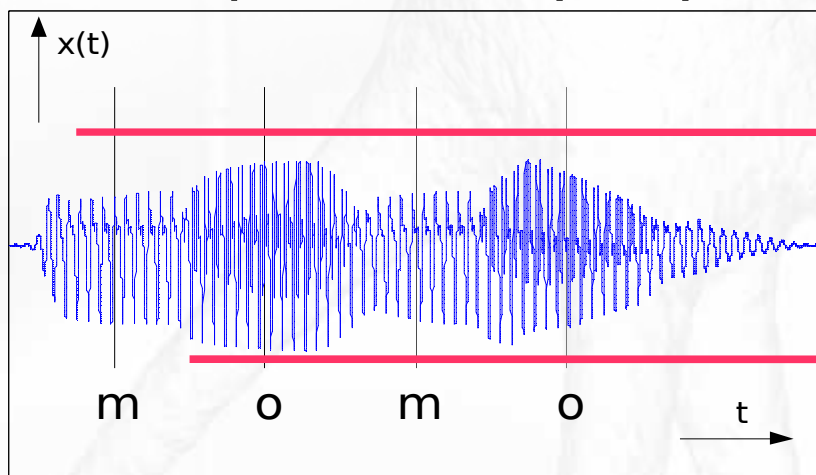
Paweł Mrówka
pawel.mrowka@neurosoft.pl

Plan prezentacji

- Synteza mowy - *SynTalk*
 - Wprowadzenie do syntezy konkatencyjnej
 - Zastosowanie analizy językowej tekstu
 - Możliwości kształtowania prozodii
- Wyszukiwanie w zasobach audio -
Videoscope
projekt
 - Cel: znajdziemy „zamrożone” informacje
 - Zastosowanie rozpoznawania

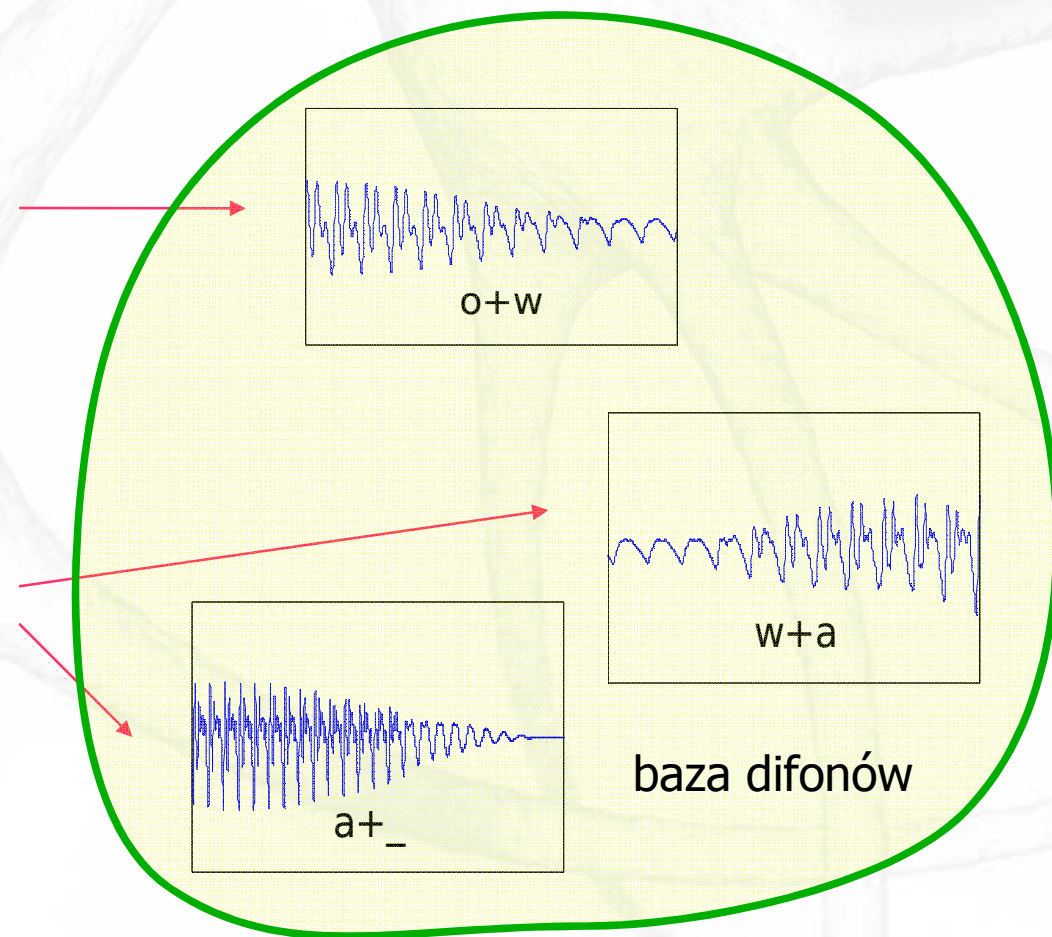
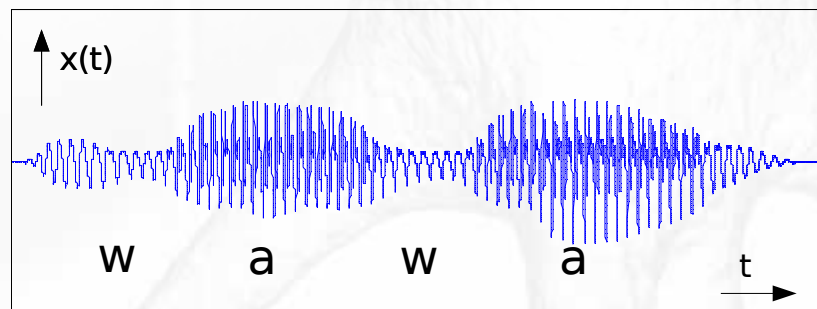
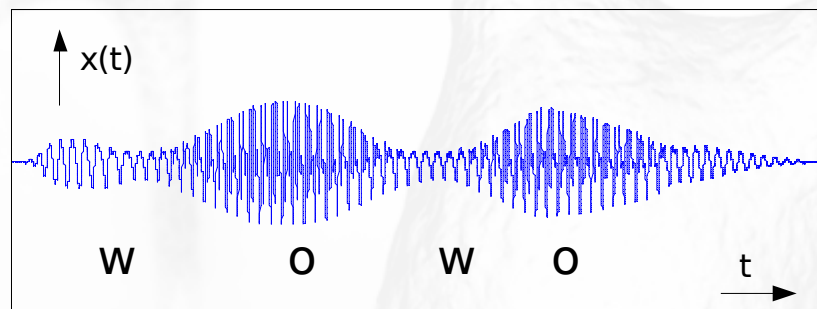
Synteza konkatencyjna

- Metoda – budowanie mowy z „cegiełek”
 - Skompletowanie bazy fragmentów mowy
 - Stosujemy *difony* (fragmenty będące przejściami między fonemami)
 - Wyodrębnienie difonów z nagrań – proces półautomatyczny



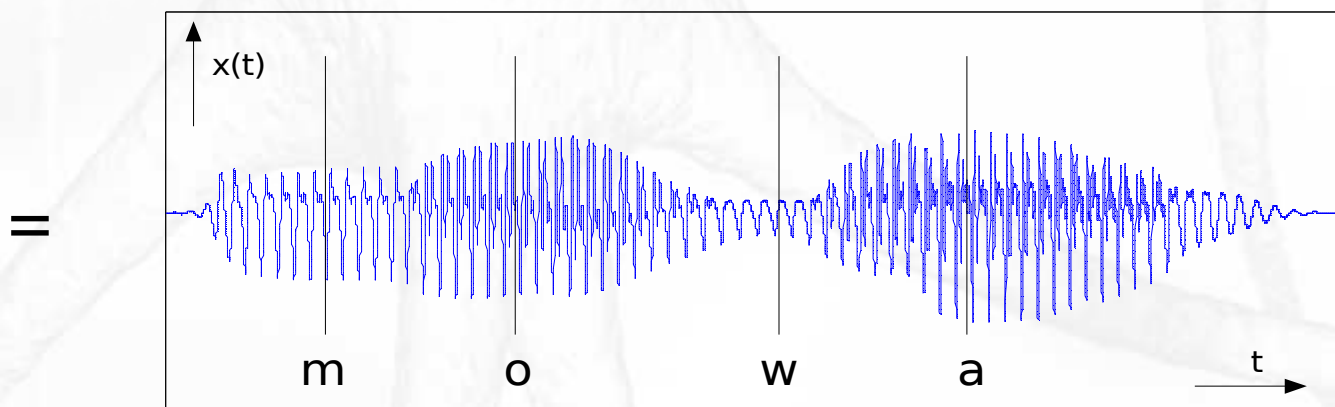
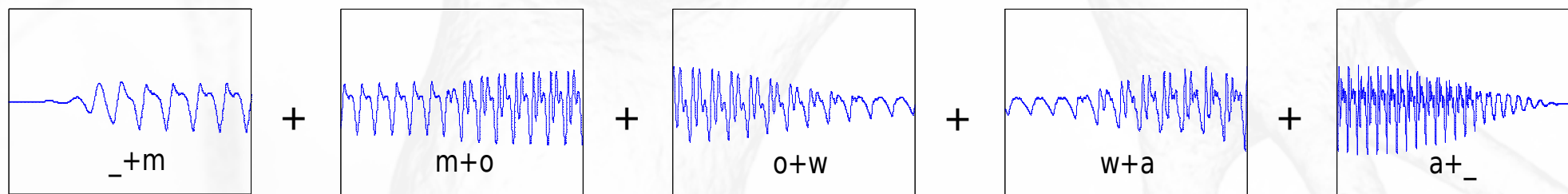
Synteza konkatenacyjna

- Skompletowanie bazy fragmentów mowy c.d.

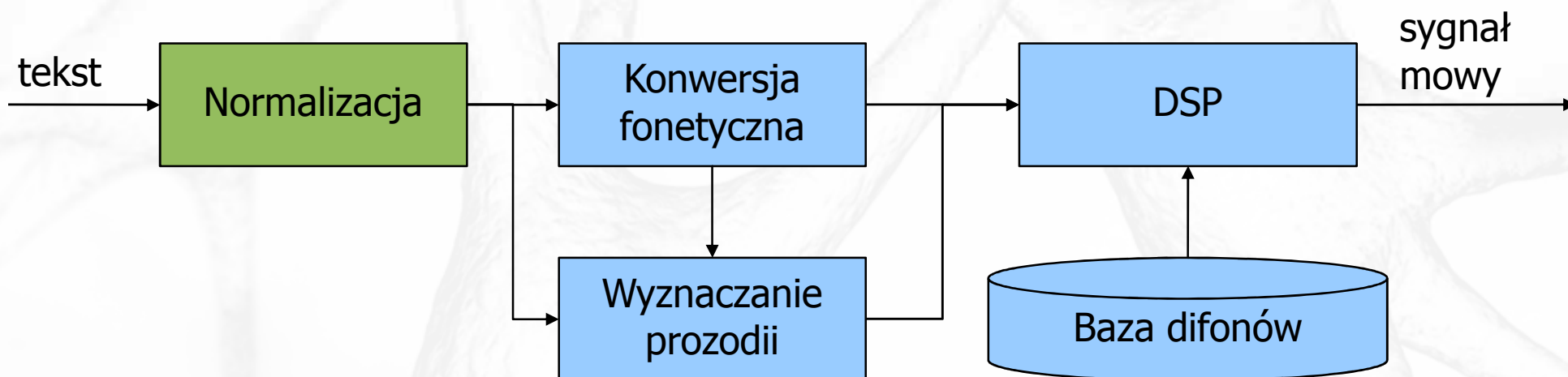


Synteza konkatenacyjna

- Łączenie difonów w zadaną wypowiedź



Schemat blokowy *SynTalka*



Normalizacja

Wejście: tekst „3 maja byliśmy w lesie.”

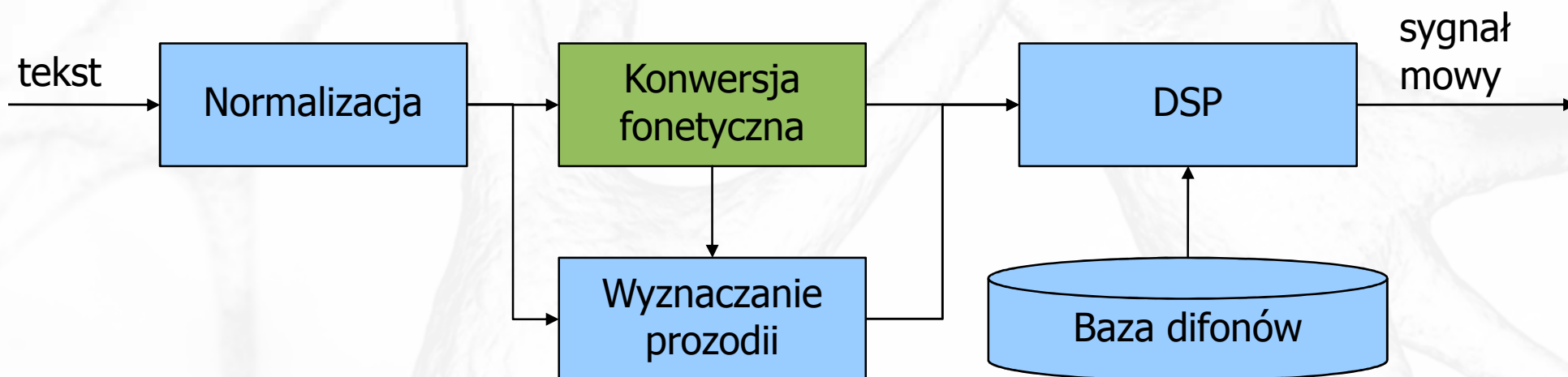
Wyjście: tekst znormalizowany + informacje dodatkowe

typ zdania: **oznajmujące**

treść: **trzeciego maja byliśmy w lesie**

ważność wyrazów: **0, +2, 0, -1, -1**

Schemat blokowy *SynTalka*



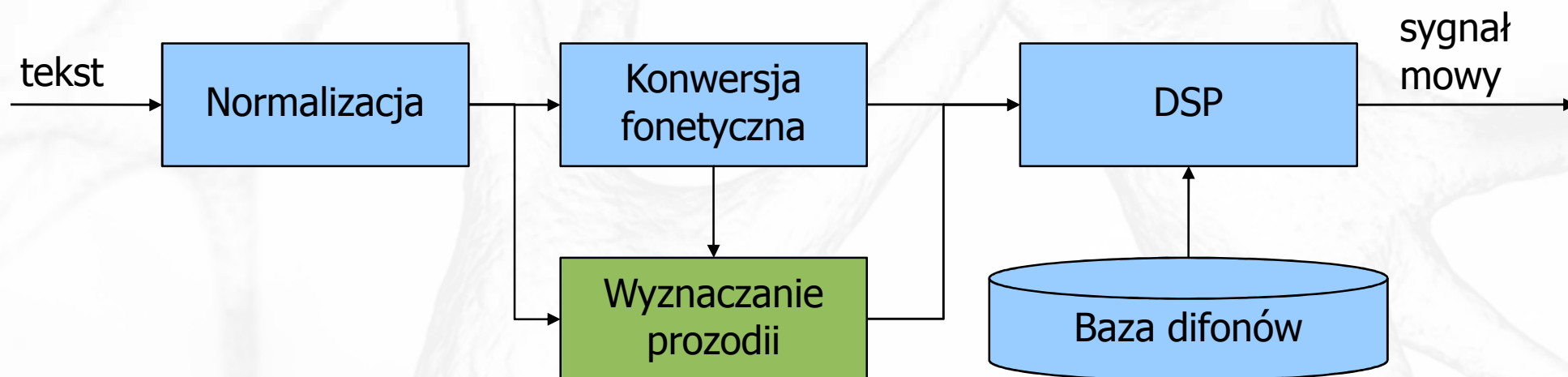
Konwersja fonetyczna

Wejście: treść zdania „trzeciego maja byliśmy w lesie”

Wyjście: ciąg difonów

**_+t t+sz sz+e e+. ci+e e+g g+o o+m m+a a+j j+a a+b
b+y y+l l+i i+si si+m m+y y+w w+l l+e e+si si+e e+_**

Schemat blokowy *SynTalka*

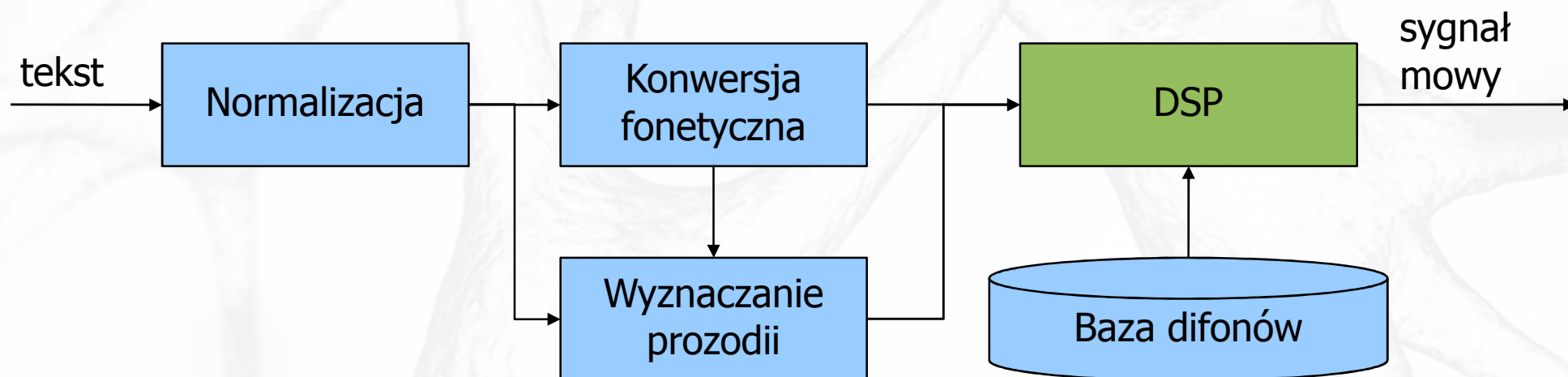


Wyznaczanie prozodii

Wejście: ciąg difonów **_+t t+sz sz+e e+. ci+e e+g g+o (...)**
 treść zdania: **trzeciego maja byliśmy w lesie**
 typ zdania: **oznajmujące**
 ważność wyrazów: **0, +2, 0, -1, -1**

Wyjście: **czasy trwania głosek** (akcenty iloczynowe)
kontur intonacyjny (akcenty toniczne, akcenty logiczne zdania)

Schemat blokowy *SynTalka*



DSP – cyfrowe przetwarzanie sygnałów

Wejście: ciąg difonów **_+t t+sz sz+e e+. ci+e e+g g+o (...)**
czasy trwania głosek
kontur intonacyjny

Wyjście: **próbki sygnału mowy** (odtworzenie, zapis do pliku...)

Analiza językowa

- Zastosowanie gramatycznej analizy tekstu w procesie normalizacji
 - Odczytywanie skrótów, liczb, dat itp.
z ich odmianą
 - Wyznaczanie logicznych akcentów wyrazowych
 - Elementy wielojęzyczności

Elastyczna prozodia

- Sterowanie przez użytkownika procesem generowania prozodii
 - Personalizacja syntezy
- Różne głosy
 - Personalizacja syntezy, czytanie z podziałem na role, synteza własnym głosem
- Transplantacja prozodii
 - Przykładowe zastosowania: książki elektroniczne, systemy dialogowe...

***SynTalk* - podsumowanie**

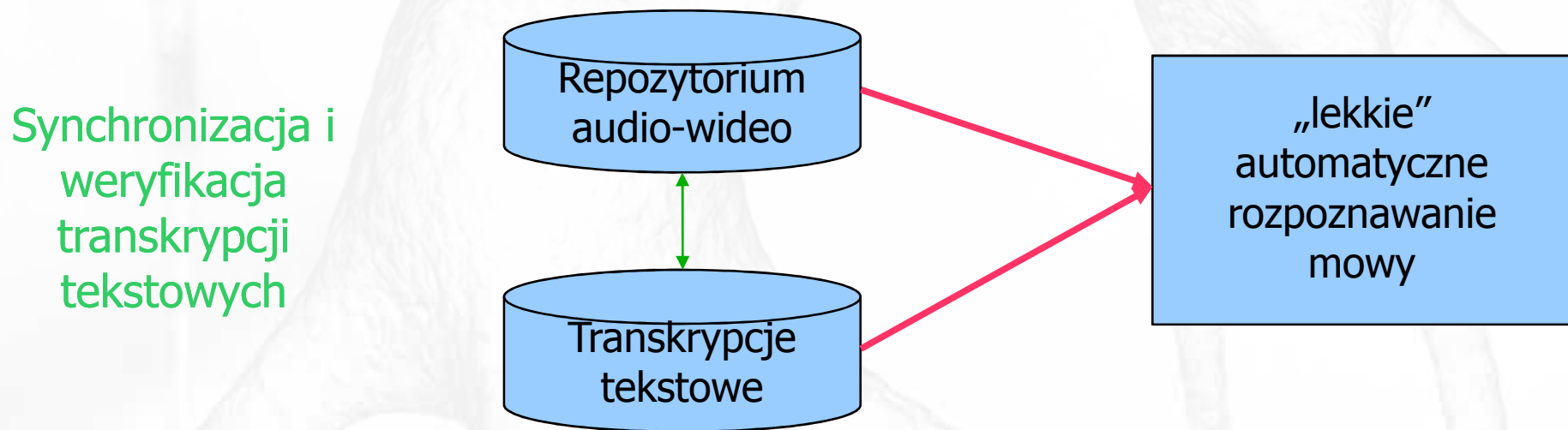
- Duże możliwości, małe wymagania:
 - Plik bazy difonów dla jednego głosu – ok. 2.5 MB
 - Zapotrzebowanie na pamięć RAM: 8 - 10 MB
 - Możliwość implementacji w urządzeniach przenośnych, systemach wbudowanych
 - Możliwość implementacji w architekturze typu serwer – lekki klient

Projekt *Videoscope*

- Cel: przeszukiwanie zasobów audiowizualnych
 - Ogromny zasób informacji – archiwalne programy radiowe, telewizyjne, dokumentacja audiowizualna obrad Sejmu...
 - Obecnie poza zasięgiem wyszukiwarek – informacja „zamrożona”

Projekt *Videoscope*

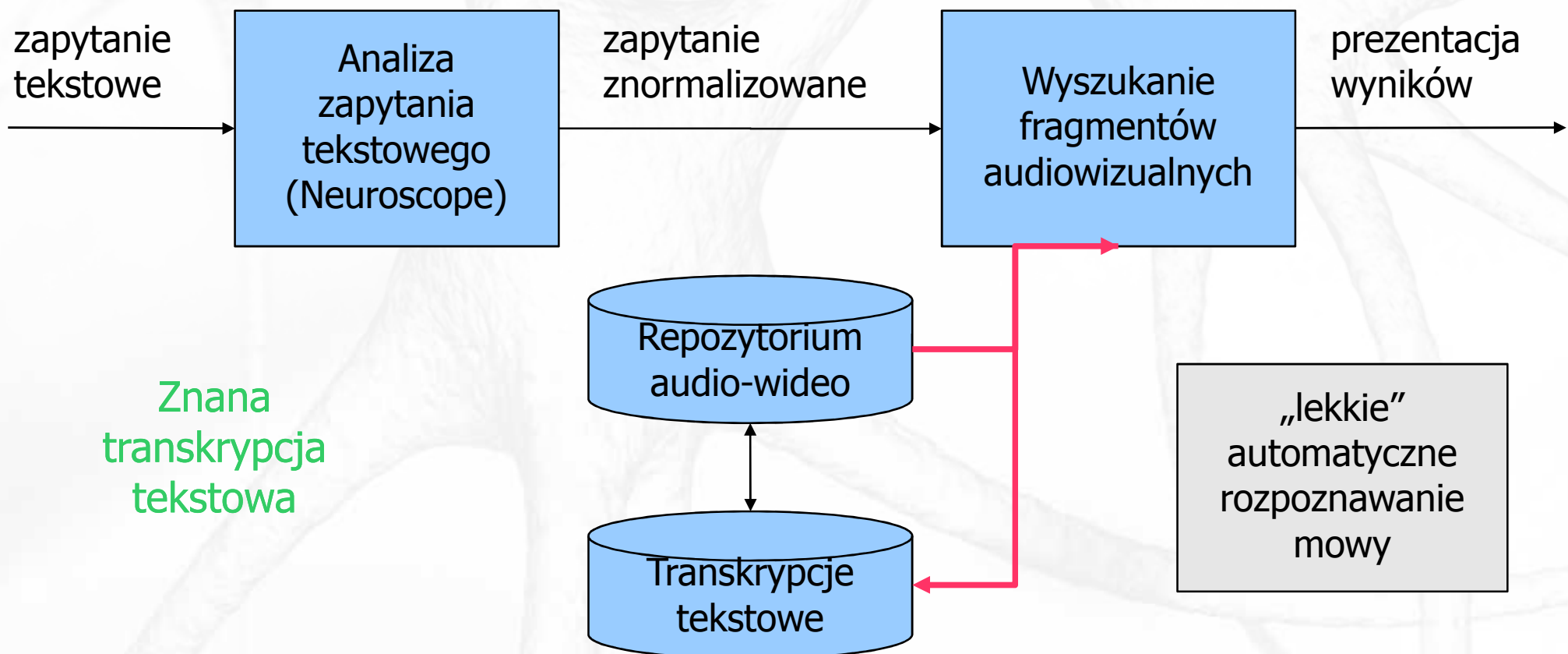
- Schemat systemu - przygotowanie



„Lekkie” rozpoznawanie mowy – wyszukiwanie fragmentów zapisu audio zawierających zadaną w postaci ortograficznej (przybliżoną) treść.

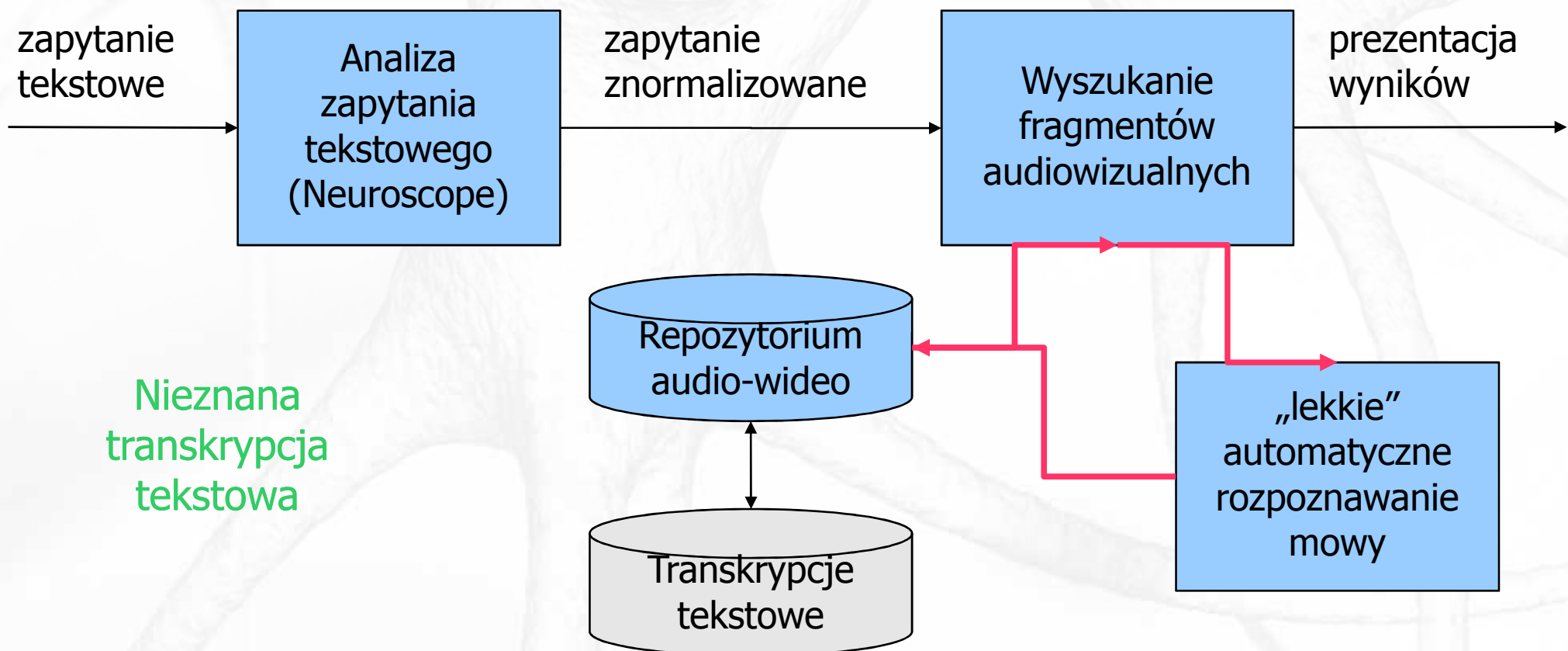
Projekt *Videoscope*

■ Schemat systemu - wyszukiwanie



Projekt *Videoscope*

■ Schemat systemu - wyszukiwanie



Projekt *Videoscope*

- Zastosowanie algorytmów rozpoznawania mowy w procesach:
 - Automatycznej weryfikacji transkrypcji tekstowej oraz jej synchronizacji czasowej z zapisem audiowizualnym
 - Wyszukiwania w zasobach audiowizualnych przy nieznanej ich transkrypcji tekstowej

Dziękuję za uwagę!