

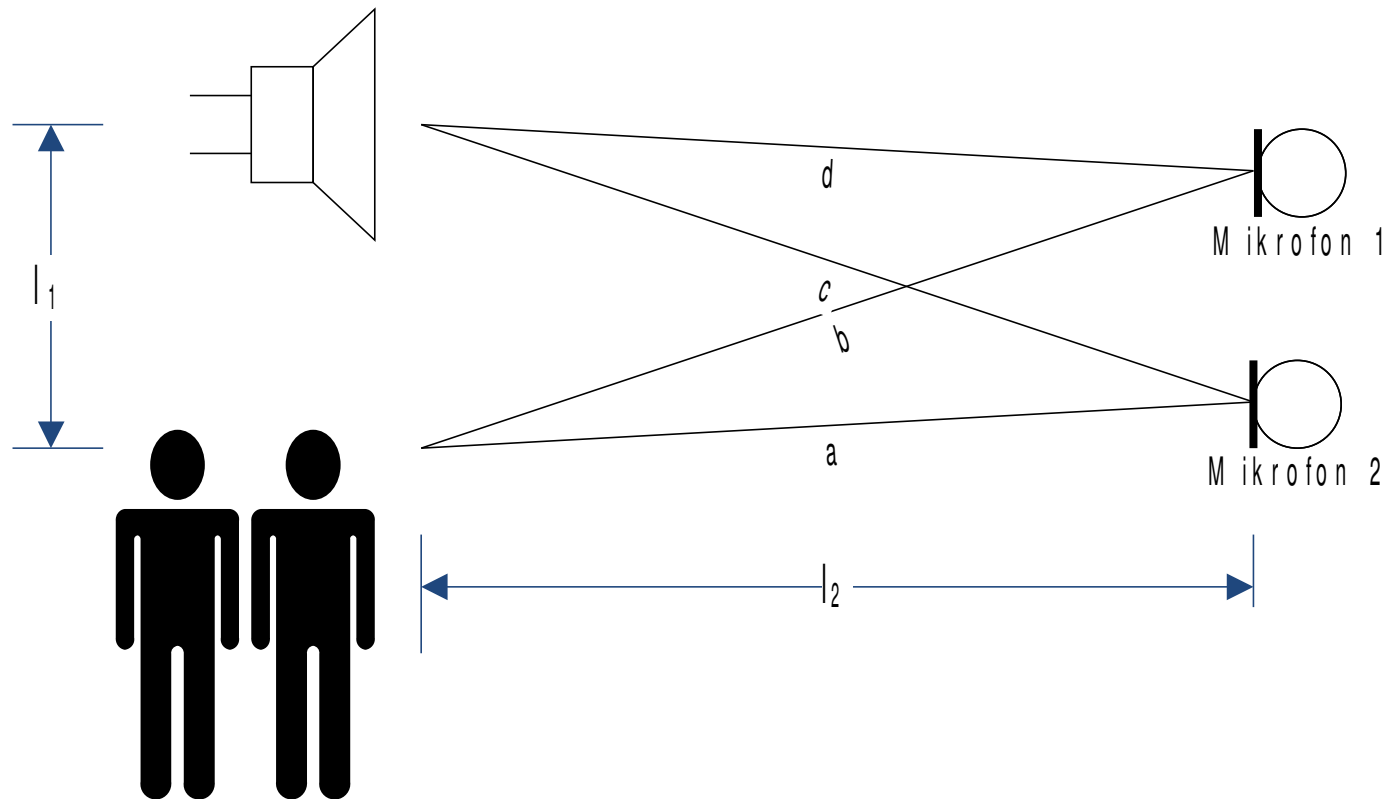
# Ekstrakcja mowy z sygnału o dużym poziomie „tła”

Rafał Samborski

# Plan prezentacji

- Zarys problemu
- Propozycja rozwiązania
- Przykład działania
- Podsumowanie

# Zarys problemu



Przy częstotliwości próbkowania 44100 Hz  
wystarczy różnica odległości na poziomie:

$$23 \mu s \cdot 330 \frac{\text{m}}{\text{s}} \approx 7,5 \text{ mm}$$

# Propozycja algorytmu

$$S_{m1}^{we} \xrightarrow{\text{BPF}} S_{m1}^{wy}$$

$$S_{m2}^{we} \xrightarrow{\text{BPF}} S_{m2}^{wy}$$

$$\tau_{corr} = \max_{\tau} \left[ \sum_n S_{m1}^{wy}(n-\tau) S_{m2}^{wy}(\tau) \right]$$

$$S_{mowa}(n) = S_{m1}^{we}(n-\tau) - k S_{m2}^{we}(n)$$

$$k = \sqrt{\frac{\sum_n S_{m1}^{wy2}(n)}{\sum_n S_{m2}^{wy2}(n)}}$$

# Postprocessing

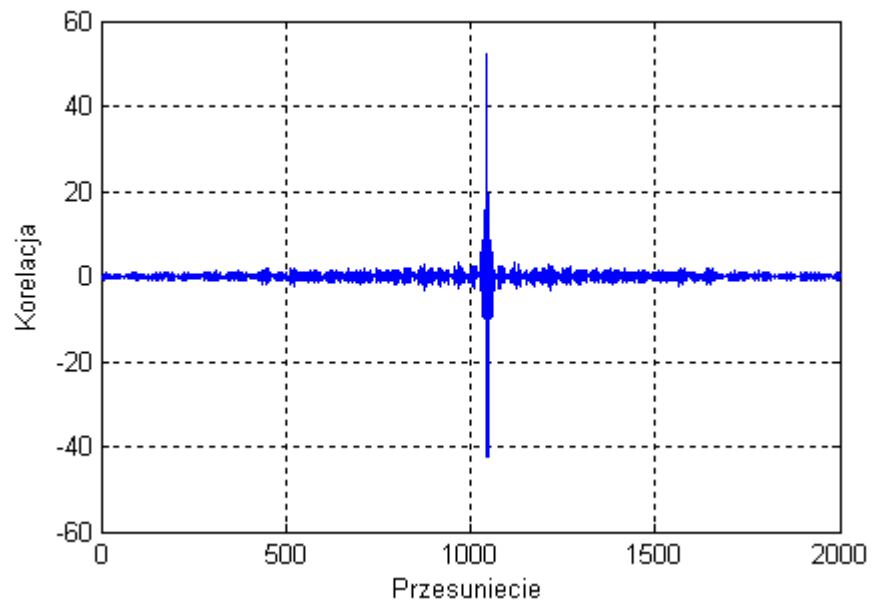
- Autokorelacja sygnału wyjściowego

$$\tau_{autocorr} = \max_{\tau} \left[ \sum_n (S_{mowa}(n - \tau)) (-S_{mowa}(\tau)) \right]$$

Kiedy jest celowe jej stosowanie i dlaczego?

- Filtracja sygnału wyjściowego

# Prezentacja działania (I)

 $S_{m1}^{we}$  $S_{m1}^{wy}$  $S_{m1}^{we}$  $S_{m2}^{wy}$ 

# Prezentacja działania (II)

$$S_{mowa}(n) = S_{m1}^{we}(n - \tau) - kS_{m2}^{we}(n)$$

# Prezentacja działania (III)

$$S_{m1}^{we}$$
$$S_{m1}^{we}$$

- Złe uwarunkowanie problemu przez zbyt czytelną warstwę mowy w nagraniach



# Podsumowanie

- Opracowany algorytm dobrze radzi sobie z sygnałem mowy ukrytym głęboko w sygnale tła
- Wprowadzenie preprocessingu – np. thresholding może zwiększyć skuteczność w przypadku bardzo zaszumionych sygnałów

# Dziękuję za uwagę

Wykorzystano nagrania:

- Richard Bona – „O Sen Sen Sen”
- Wojciech Mann – „Pan Twardowski”