Specjalne metody przetwarzania dźwięku
Filtracja przestrzenna (beamforming)
Rozpoznawanie sygnałów fonicznych

Józef Kotus

12.12.2023

Plan wykładu

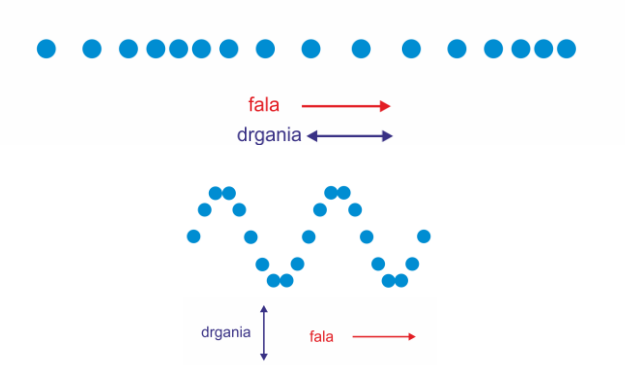
- Wprowadzenie
- Beamforming
- Przykłady zastosowań

Wprowadzenie

- Przydatne, podstawowe pojęcia
  - Fala akustyczna
  - Ciśnienie akustyczne
  - Prędkość akustyczna
  - Natężenie dźwięku
  - Pole swobodne
  - Pole dyfuzyjne
  - Przetworniki elektroakustyczne
  - Charakterystyka kierunkowa

Fala akustyczna

- Fala akustyczna (dźwiękowa) jest jedną z form przenoszenia energii. Polega ona na cyklicznym przemieszczaniu się cząsteczek sprężystego środowiska wokół położenia równowagi – tzw. drgania mechaniczne.
- Podstawowy podział to:
  - fale podłużne - kierunek przemieszczeń cząsteczek pokrywa się z kierunkiem przenoszenia energii;
  - fale poprzeczne - kierunek przemieszczeń cząsteczek jest prostopadły do kierunku przenoszenia energii.

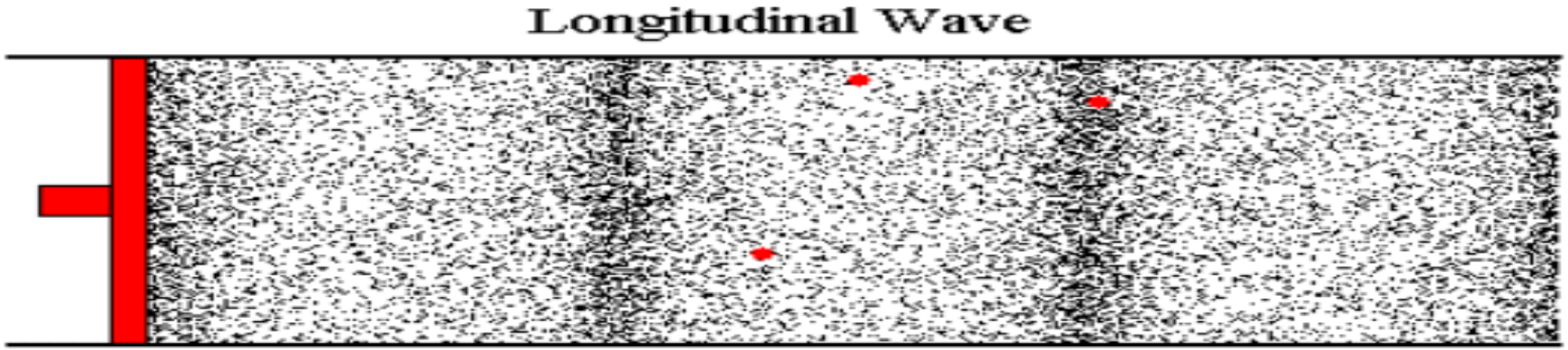


- Wewnątrz ośrodków nie posiadających sprężystości postaci (gazy i ciecze, które przyjmują kształt zawierającego je zbiornika) fala akustyczna jest wyłącznie falą podłużną.
- W ciałach stałych występują zarówno fale podłużne jak i poprzeczne, a także inne, o bardziej złożonym charakterze

https://home.agh.edu.pl/~kakol/efizyka/w13/main13a.html

Ciśnienie akustyczne, prędkość akustyczna, natężenie fali

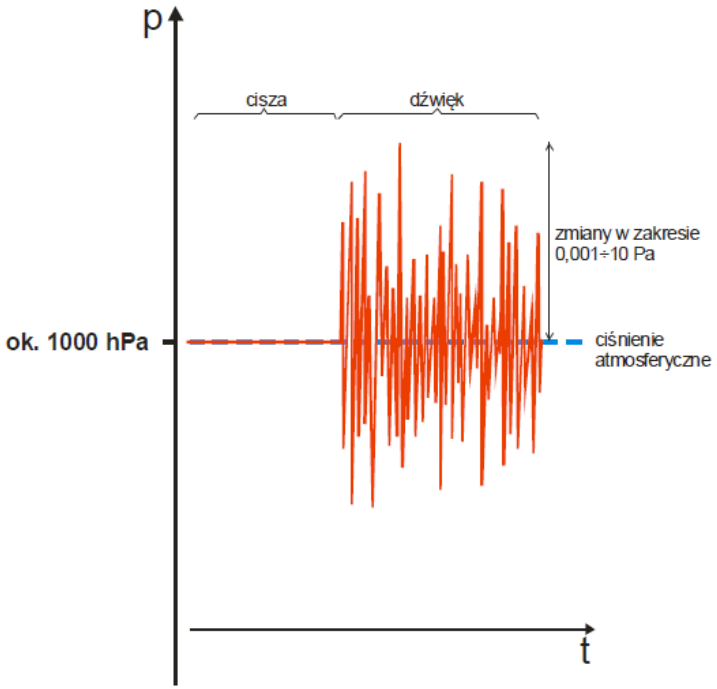
- Pod pojęciem ciśnienia akustycznego należy rozumieć zmienną w czasie nadwyżkę ciśnienia (ponad ciśnienie atmosferyczne) wywołaną obecnością fali akustycznej.
- Pod pojęciem prędkości akustycznej należy rozumień prędkość cząstki akustycznej w jej ruchu drgającym.
- W przypadku fal akustycznych natężenie fali definiowane jest jako iloczyn ciśnienia akustycznego przez prędkość cząstki akustycznej w jej ruchu drgającym.

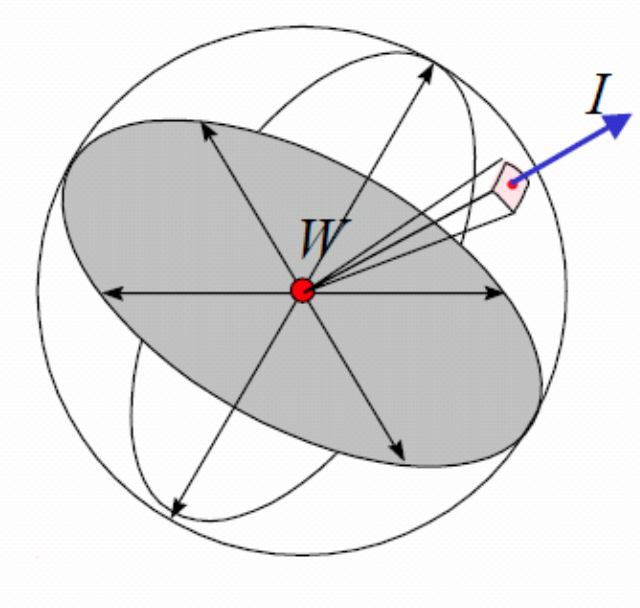


Ciśnienie akustyczne, prędkość akustyczna, natężenie fali

- **Ciśnienie akustyczne**
  - chwilowe zmiany ciśnienia względem średniego ciśnienia atmosferycznego:

$$I = p \cdot v,$$





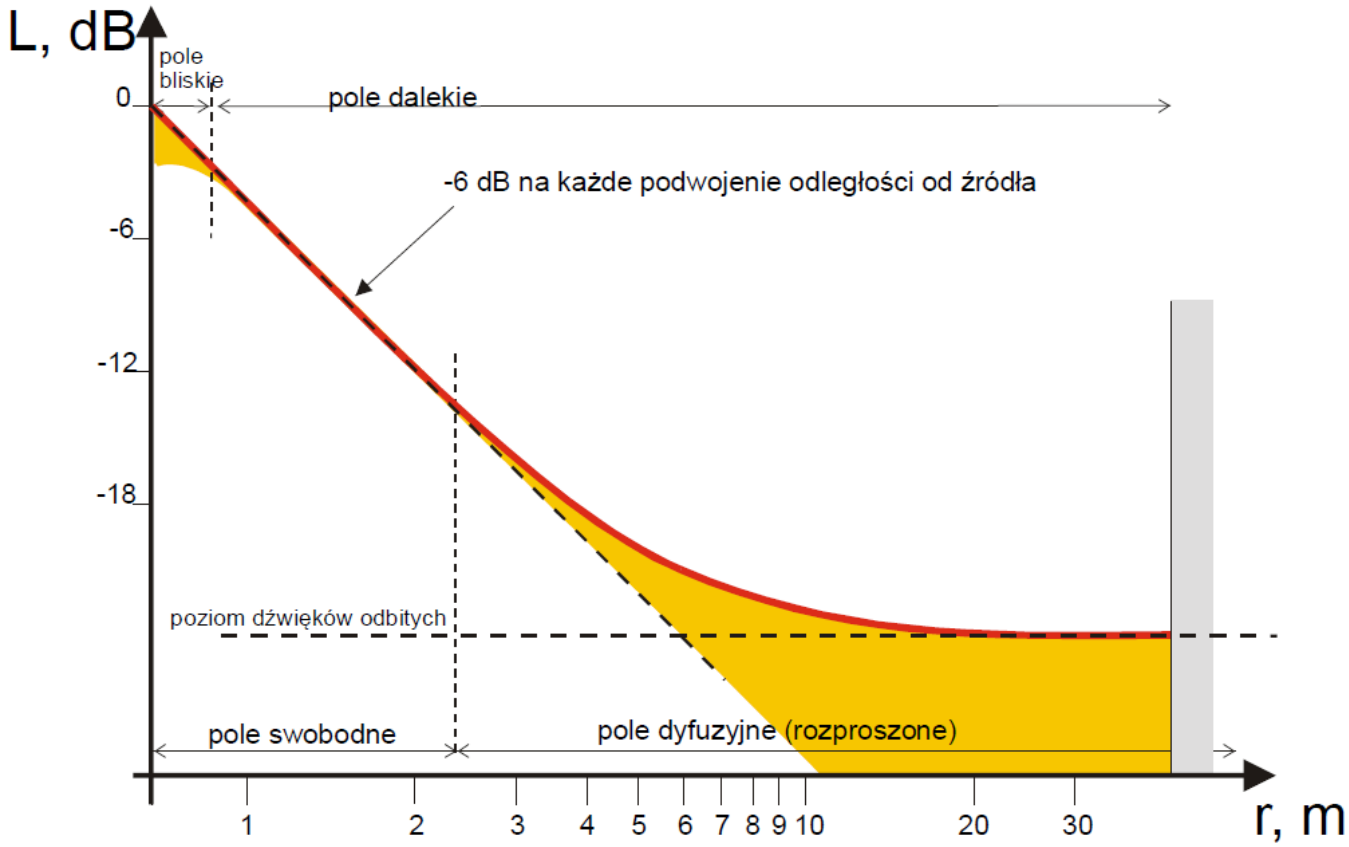
Ciekawostka (definicje będą dalej...):
*- minimalny poziom dźwięku to **-∞** [dB] - czyli brak emisji, ale maksymalny poziom (fala sinusoidalna, w powietrzu, dla warunków normalnych t=0°C, $p_{atm.}$=101325 Pa) - to tylko **194,1 dB**...*

$$L_p = 10 \cdot \lg \frac{p^2}{p_0^2},\ dB$$

gdzie: p_0 - ciśnienie odniesienia $2*10^{-5}$ Pa (próg słyszenia dla 1000 Hz)

Klasyfikacja pól akustycznych

- Pole akustyczne:
  - **bliskie** - obszar pola bezpośrednio przylegający do źródła dźwięku, gdzie występują zjawiska nieliniowe (około 1 długości fali - dla 250 Hz jest to 2,5 m !!!).
  - **dalekie** - obszar pola, w którym spadek poziomu dźwięku wynosi 6 dB na każde podwojenie odległości od źródła hałasu (dla fali sferycznej - od źródła punktowego)
  - **swobodne** - pole w którym nie występują fale odbite,
  - **dyfuzyjne** - pole w którym występuje duża liczba fal odbitych z różnych kierunków, co powoduje względnie stały poziom dźwięku w całym obszarze

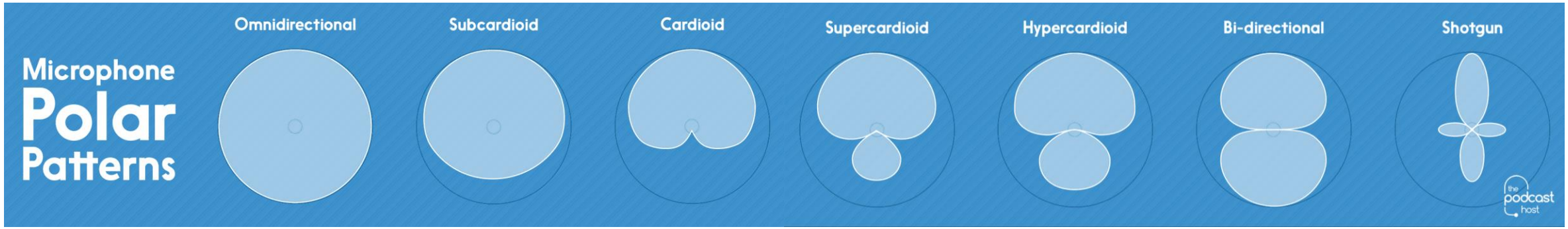


Przetworniki elektroakustyczne

- Przetwornik elektroakustyczny – urządzenie przetwarzające prąd elektryczny na fale akustyczne lub odwrotnie.
- Do przetworników elektroakustycznych zalicza się m.in głośniki, mikrofony, słuchawki, geofony i hydrofony.
- Podział przetworników ze względu na wykorzystywane zjawisko fizyczne:
  - elektromagnetyczne,
  - magnetoelektryczne,
  - magnetostrykcyjne,
  - pojemnościowe,
  - elektrostatyczne,
  - piezoelektryczne.

Charakterystyka kierunkowa (np. mikrofonu)

- Właściwości kierunkowe mikrofonu określone są stosunkiem skuteczności przy dowolnym kierunku padania fali dźwiękowej na mikrofon do skuteczności przy padaniu prostopadłym na element odbierający energię akustyczną. Przebieg tego stosunku w funkcji kąta padania fali nazywa się charakterystyką kierunkową
- Charakterystyka kierunkowości mikrofonu określa jego czułość na dźwięki w odniesieniu do kierunku lub kąta, pod którym te dźwięki docierają. Inaczej mówiąc jest to sposób, w jaki mikrofon „słyszy” dźwięki dochodzące z różnych kierunków.



Beamforming

Beamforming - wprowadzenie

- Beamforming - kształtowanie wiązki
  - Technika przetwarzania sygnału wykorzystywana w macierzach sensorowych do kierunkowej transmisji oraz odbioru sygnału.
  - Efekt jest uzyskiwany poprzez ułożenie elementów promieniujących w pewien szyk w taki sposób, że sygnały pod pewnymi kątami wzmacniają się, a pod innymi wytłumiają.
  - Kształtowanie wiązki może być wykorzystywane zarówno po stronie nadawczej, jak i odbiorczej w celu osiągnięcia selektywności przestrzennej.
  - Poprawa w porównaniu z odbiorem/nadawaniem dookolnym jest rozumiana jako zysk (lub strata).
  - Kształtowanie wiązki może być wykorzystywane dla fal radiowych, jak i akustycznych.
  - Technika ta ma wiele zastosowań: w radarach, sonarach, sejsmologii, komunikacji bezprzewodowej, radioastronomii, mowie, akustyce, biomedycynie….

Filtracja przestrzenna

- Filtracja przestrzenna jest przekształceniem sygnału uwzględniającym wzajemne usytuowanie źródła i odbiornika
- Celem filtracji przestrzennej jest **kształtowanie** charakterystyki kierunkowej przetwornika dźwięku
- Wykorzystanie algorytmów cyfrowego przetwarzania sygnałów do uzyskania pożądanej charakterystyki
- Wykonywana jest najczęściej po stronie odbiorczej

Beamforming - wprowadzenie

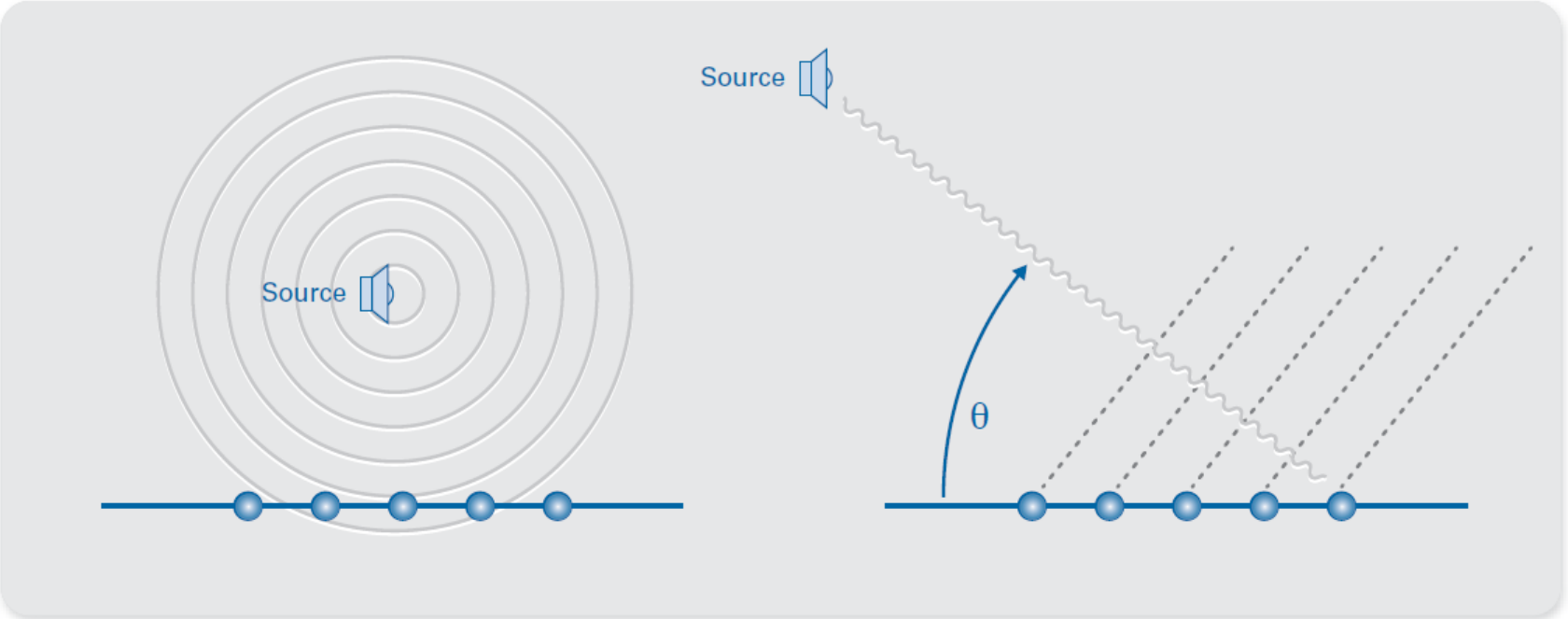


Figure 1. Sound waves in the near field (A) move away from the source in a circular pattern. In the far field (B), they are far enough away from the source that they appear to move in a straight line.

$$spatial\ resolution = \frac{distance\ from\ source}{diameter\ of\ the\ array} \times sound\ wavelength$$

Beamforming - wprowadzenie

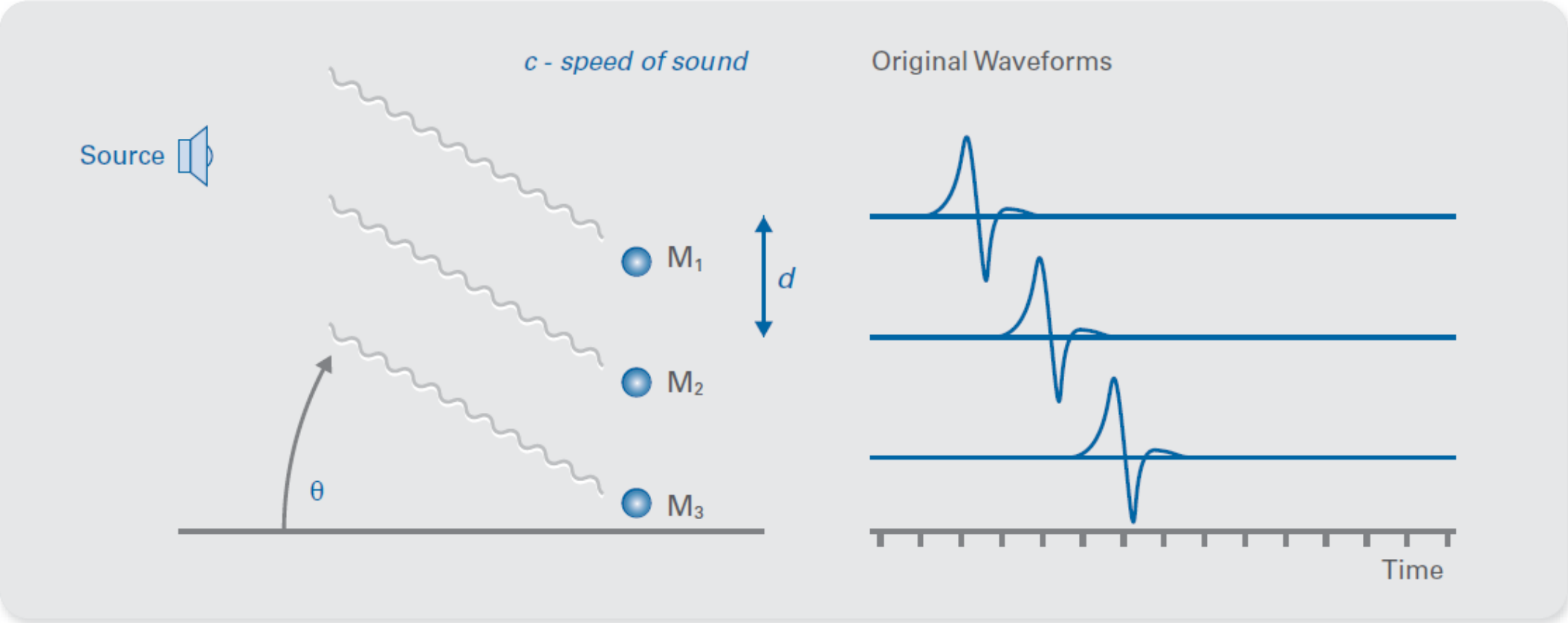


Figure 8. If the noise source is offset from the microphone array, the sound waves reach the closest microphones first, causing a measurable time delay between each microphone based on the distance between them.

Beamforming - wprowadzenie

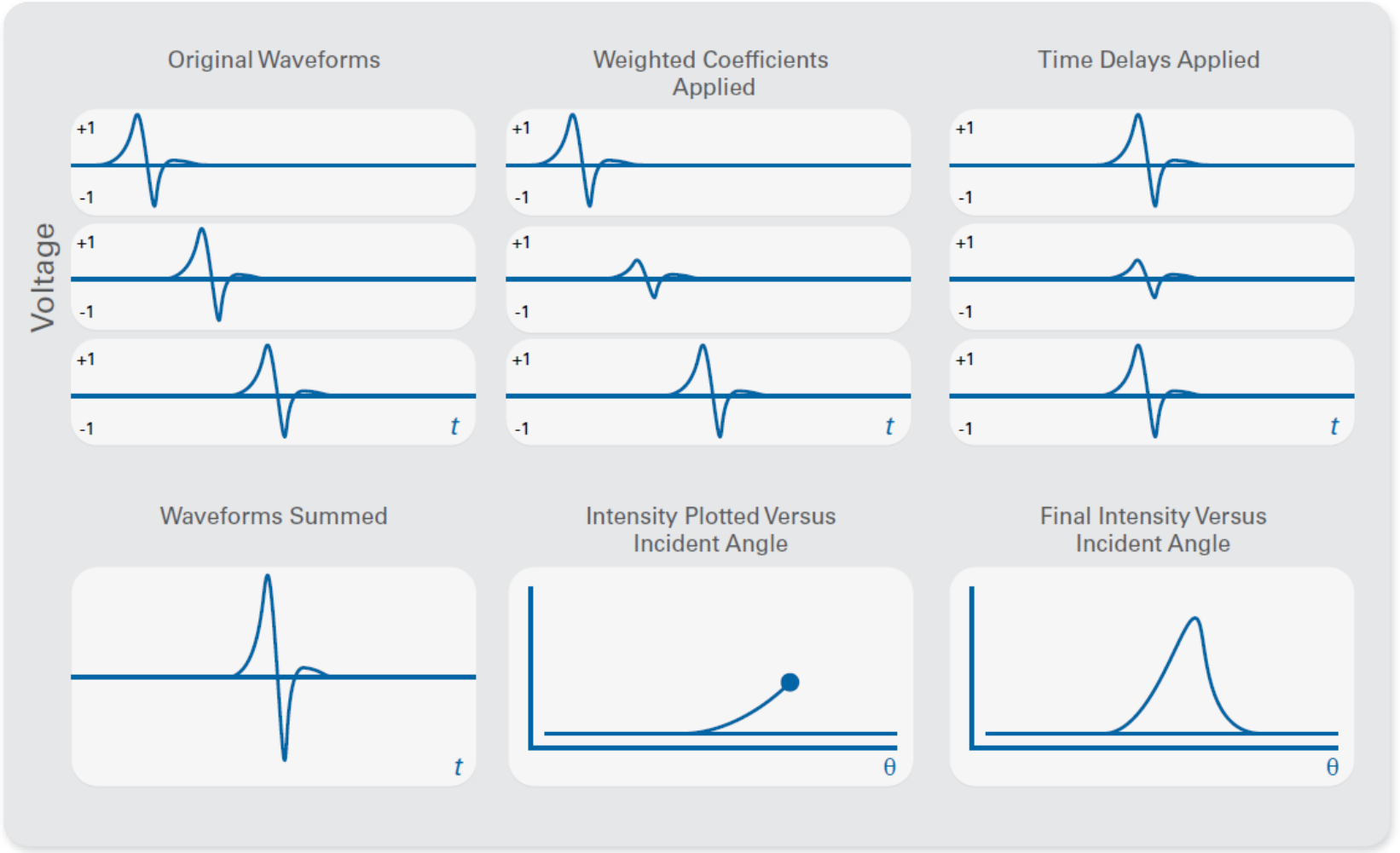
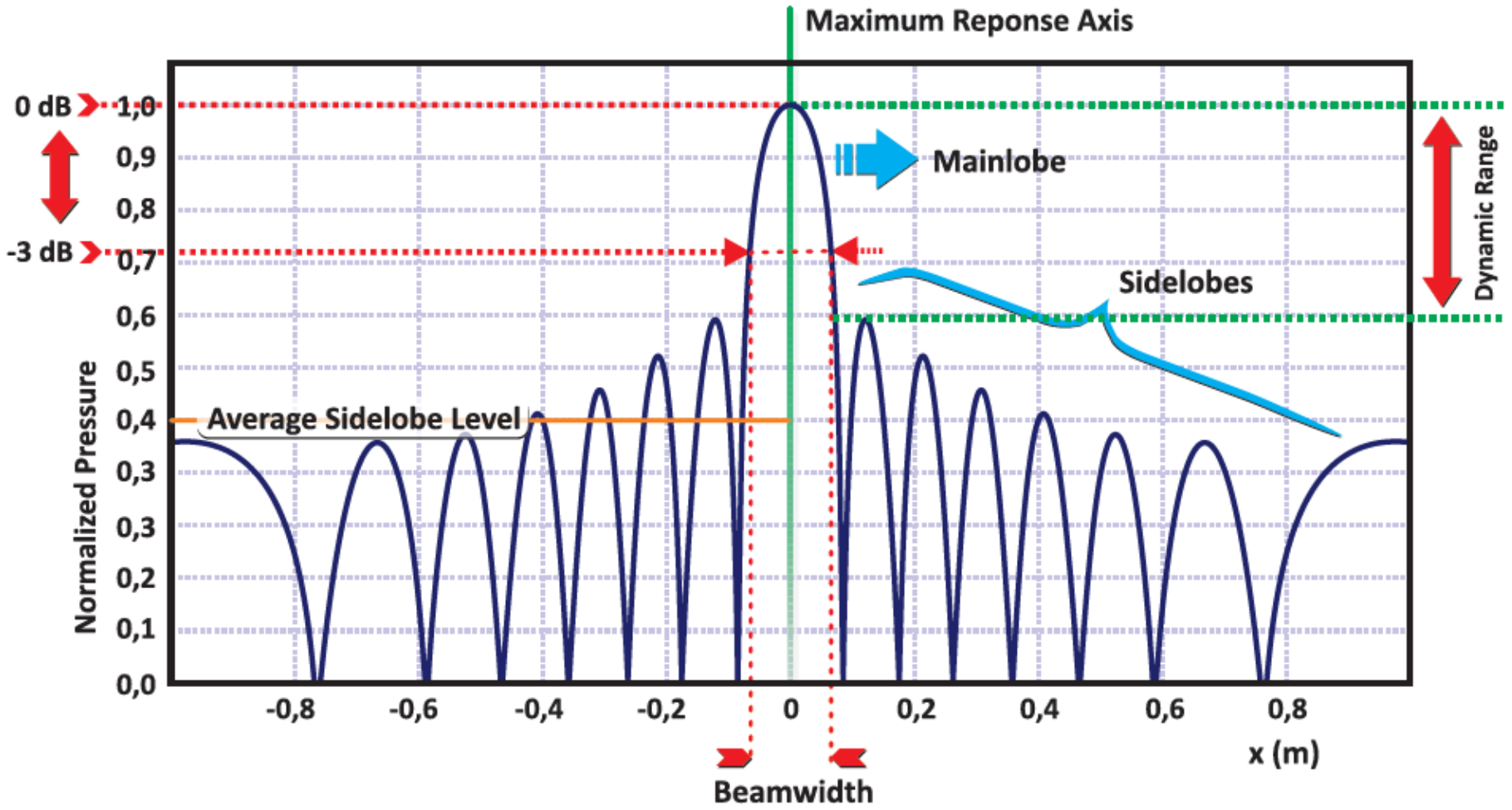


Figure 9. The delay and sum method is one of the simplest and most common algorithms used for acoustic beamforming.

Beamforming - wprowadzenie



Beamforming – podstawowe techniki

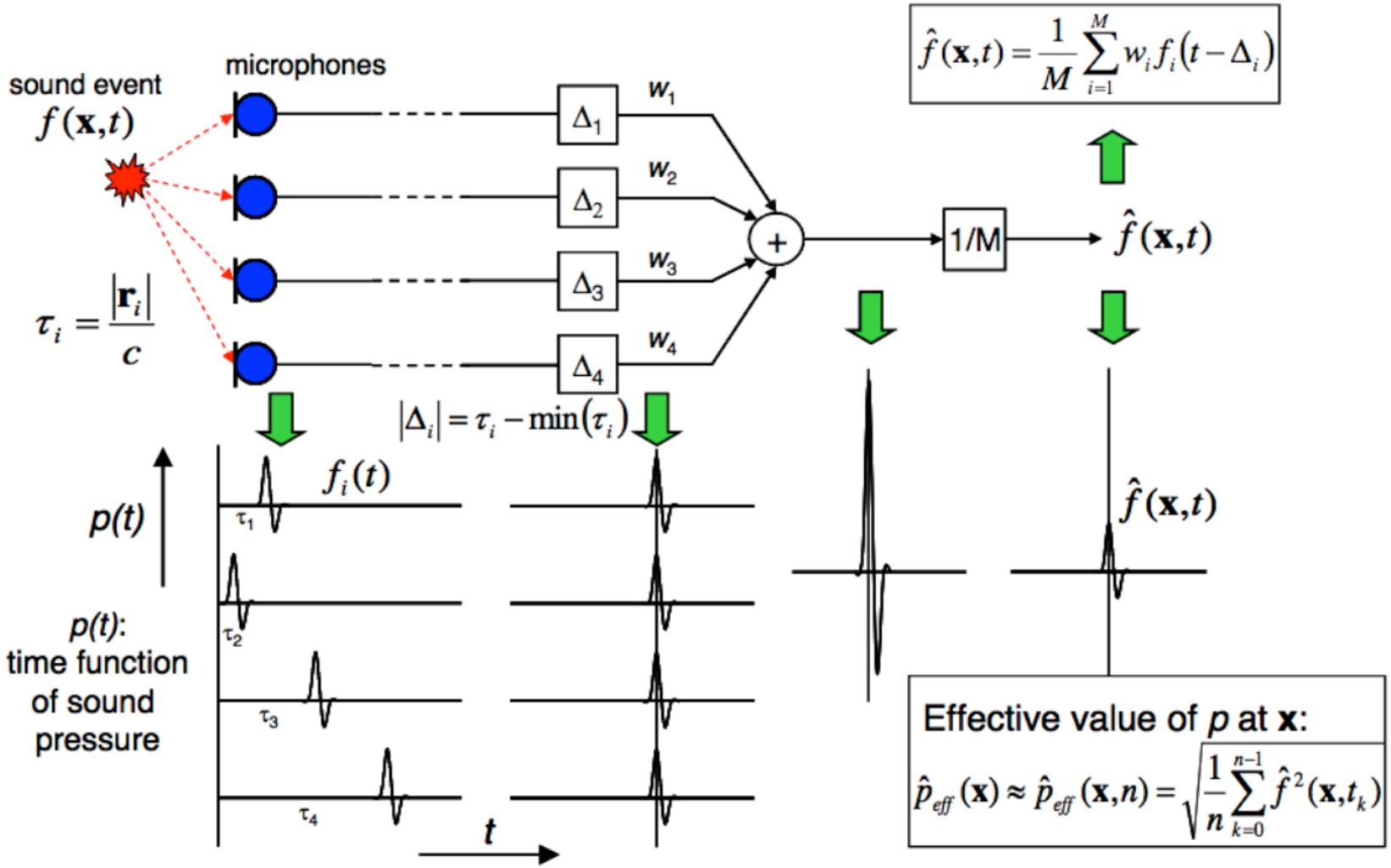
Podział ze względu na dziedzinę przetwarzania sygnału

- Metody działające w dziedzinie czasu
- Metody działające w dziedzinie częstotliwości

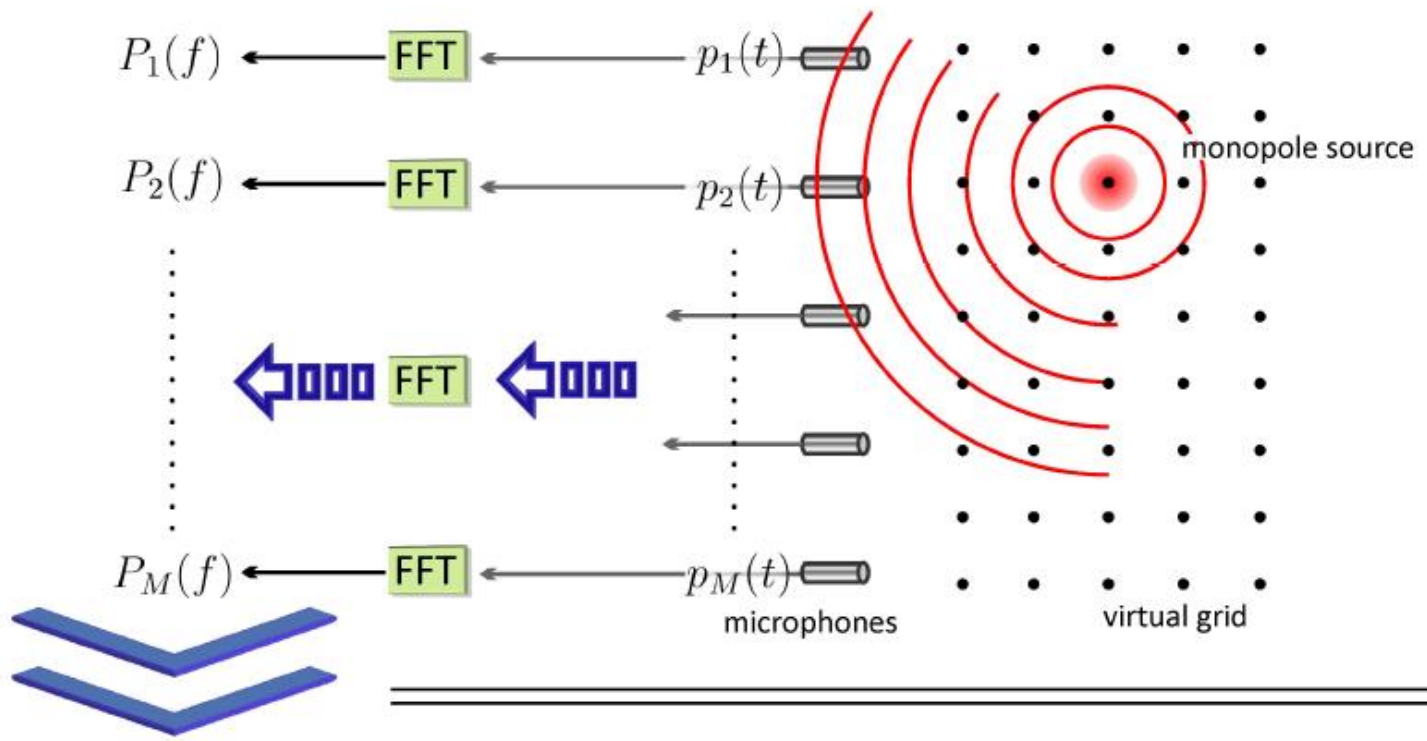
- Podział ze względu na rozmiary macierzy przetworników
  - Rozległe
  - Skupione (wektorowy czujnik akustyczny)
- Podział ze względu na rodzaj czujnika akustycznego
  - Mikrofonowe (wrażliwych na ciśnienie akustyczne)
  - Z wykorzystaniem mikroprzepływomierzy (wrażliwych na prędkość akustyczną)
- Podział ze względu na typ przetworników
  - Po stronie odbiorczej (zastosowanie mikrofonów)
  - Po stronie nadawczej (zastosowanie głośników) – synteza pola akustycznego

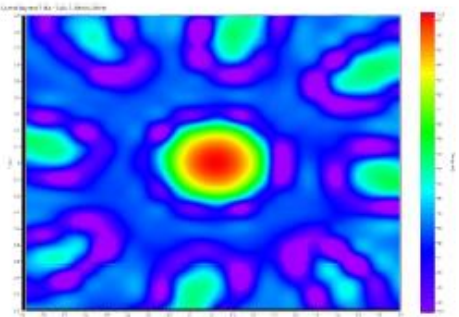
Beamforming po stronie odbiorczej

Beamforming – dziedzina czasu



Beamforming – dziedzina częstotliwości





Array Pressure Response
Output of Beamforming

$$\sqrt{|B_k|^2}$$

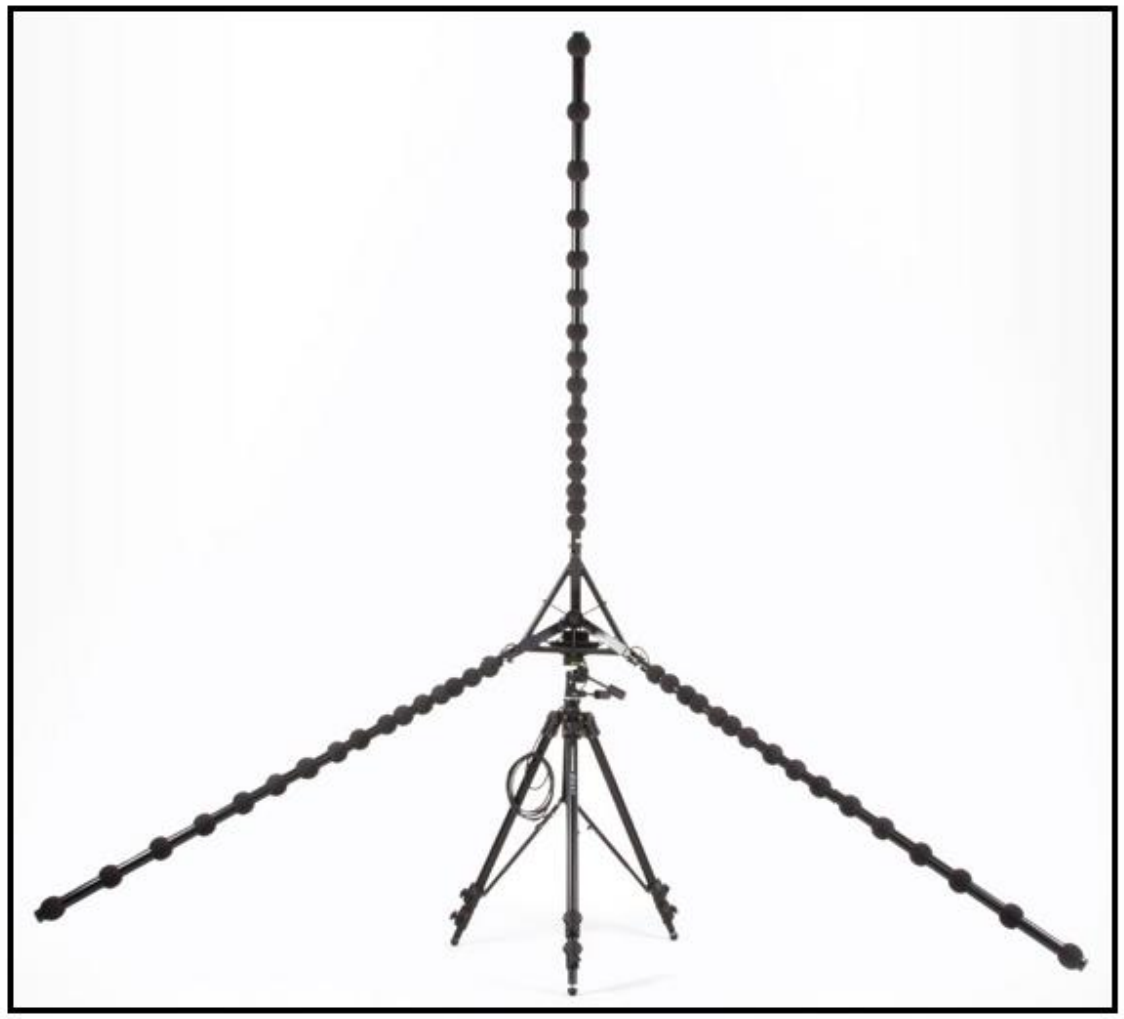
Cross Spectral Matrix (CSM)

Steering Vector

Array Power Response

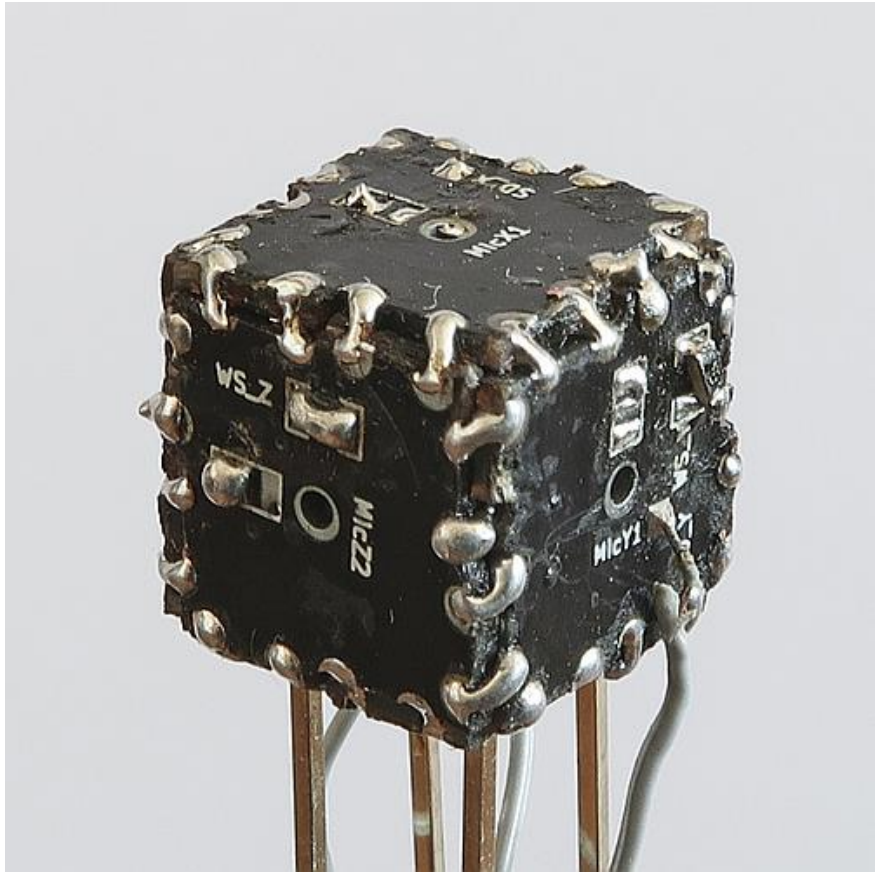
$$C_k = \begin{bmatrix} C_{11_k} & C_{12_k} & \dots & C_{1M_k} \\ C_{21_k} & C_{22_k} & \dots & C_{2M_k} \\ \vdots & \vdots & \ddots & \vdots \\ C_{M1_k} & C_{M2_k} & \dots & C_{MM_k} \end{bmatrix} \Rightarrow \vec{g_k} = \begin{bmatrix} w_1 e^{j\omega_k \Delta_1} \\ \vdots \\ w_M e^{j\omega_k \Delta_M} \end{bmatrix} \Rightarrow \begin{bmatrix} w_1 e^{j\omega_k \Delta_1} \\ \vdots \\ w_M e^{j\omega_k \Delta_M} \end{bmatrix}^H \cdot \left[\frac{1}{M^2}\right] \cdot \begin{bmatrix} C_{11_k} & C_{12_k} & \dots & C_{1M_k} \\ C_{21_k} & C_{22_k} & \dots & C_{2M_k} \\ \vdots & \vdots & \ddots & \vdots \\ C_{M1_k} & C_{M2_k} & \dots & C_{MM_k} \end{bmatrix} \cdot \begin{bmatrix} w_1 e^{j\omega_k \Delta_1} \\ \vdots \\ w_M e^{j\omega_k \Delta_M} \end{bmatrix}$$

Beamforming – matryce rozległe (duże rozmiary)



Beamforming – matryca skupiona (małe wymiary)

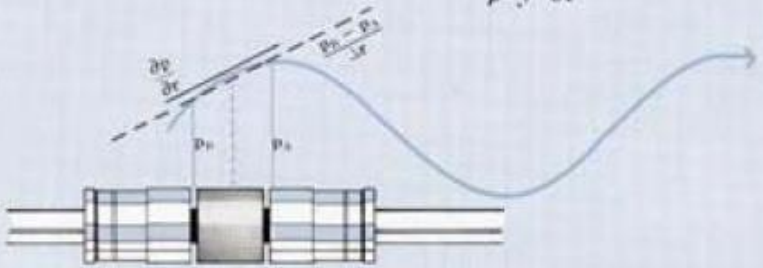
- Matryca ma wymiary 1 cm3, składa się z 6 mikrofonów typu MEMS, przykład wektorowego czujnika akustycznego typu pp (ciśnienie - ciśnienie)



Time Domain Formulation

From Euler

$$u = -\frac{1}{\rho}\int \frac{\partial p}{\partial r}\,dt$$



The Finite Difference Approximation

$$u = -\frac{1}{\rho}\int \frac{p_B - p_A}{\Delta r}\,dt$$

Average pressure

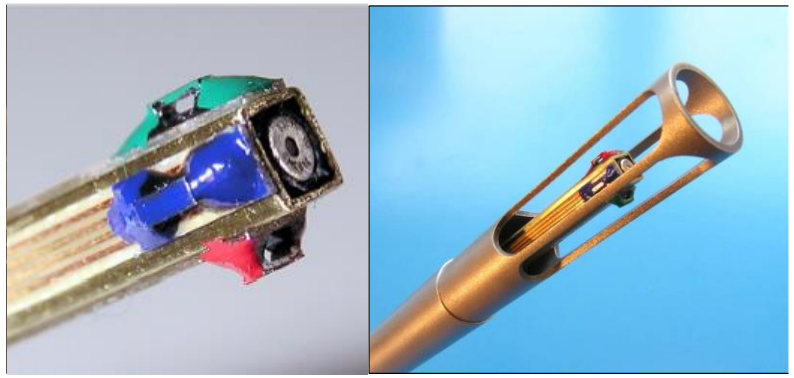
$$p = \frac{p_A + p_B}{2}$$

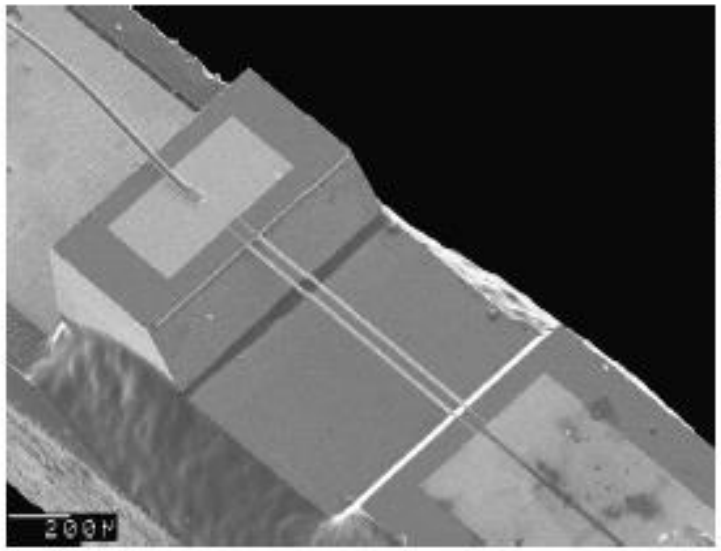
$$I = \overline{p \cdot u}$$

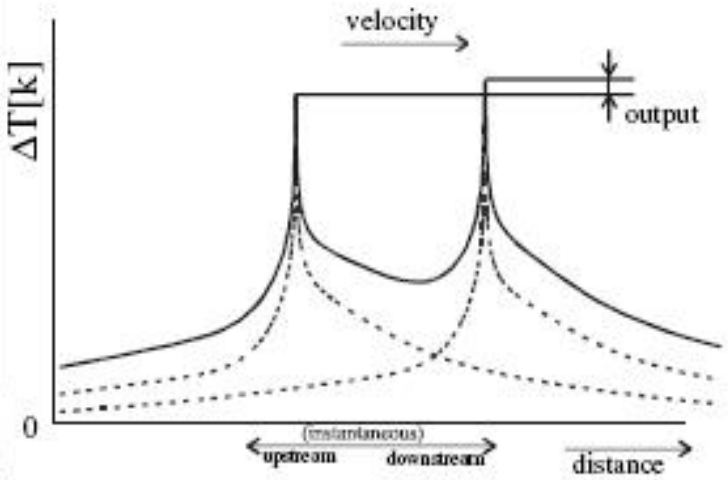
$$I = -\frac{p_A + p_B}{2\rho\Delta r}\int (p_B - p_A)\,dt$$

Beamforming – matryca skupiona (małe wymiary)

- Przykład wektorowego czujnika akustycznego typu pu (ciśnienie - prędkość)
- Czujnik wrażliwy na prędkość akustyczną to mikroprzepływomierz cząsteczkowy (microflown)







Przykłady zastosowań

Bemforming – przykładowe zastosowania

- Po stronie odbiorczej
  - Lokalizacja źródeł dźwięku
  - Obrazowanie pola akustycznego
  - Nieinwazyjna diagnostyka obiektów
- Po stronie nadawczej
  - Synteza pola akustycznego
  - Kształtowanie akustyki wnętrz (kontrola czasu pogłosu)

Beanforming – przykład dla źródeł w ruchu

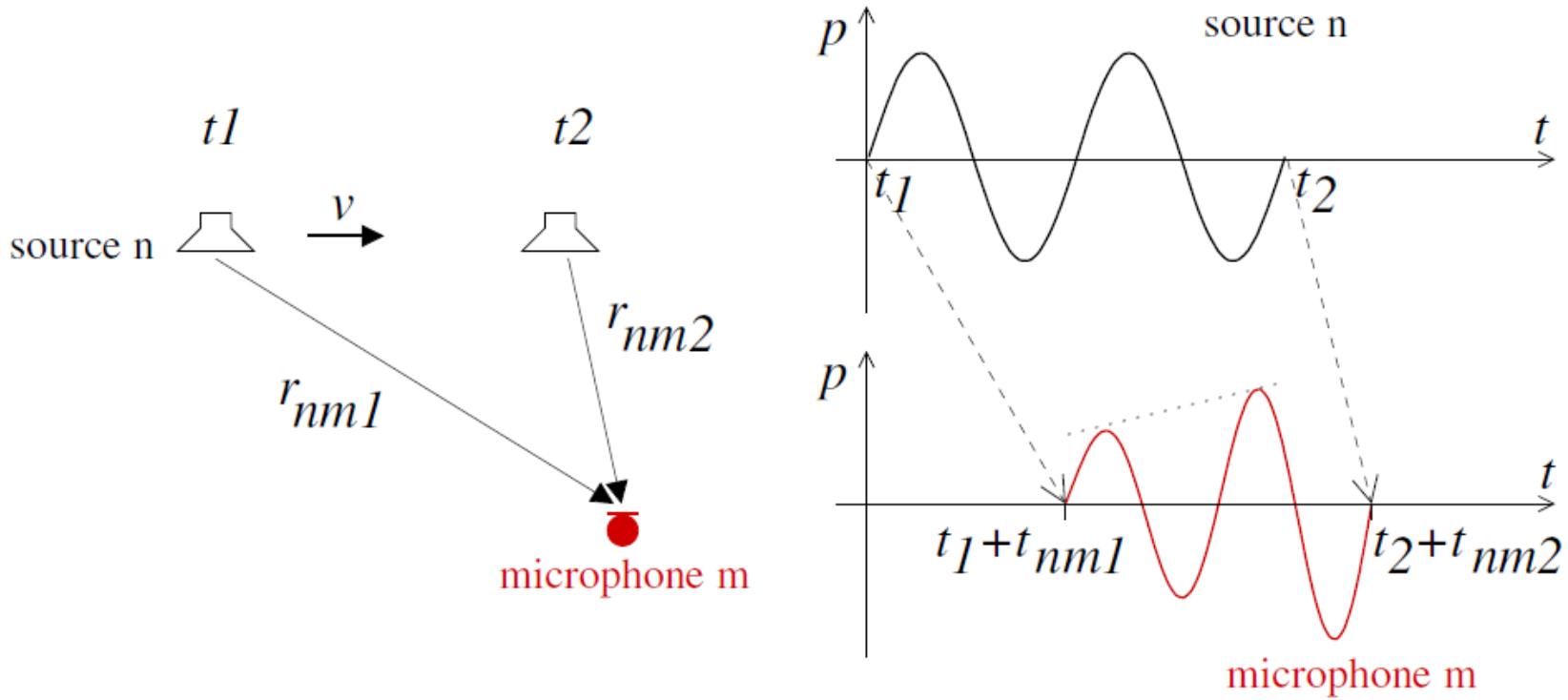


Figure 1: Derivation of the Doppler effect

Beanforming – przykład dla źródeł w ruchu

- Macierz 32 mikrofonów

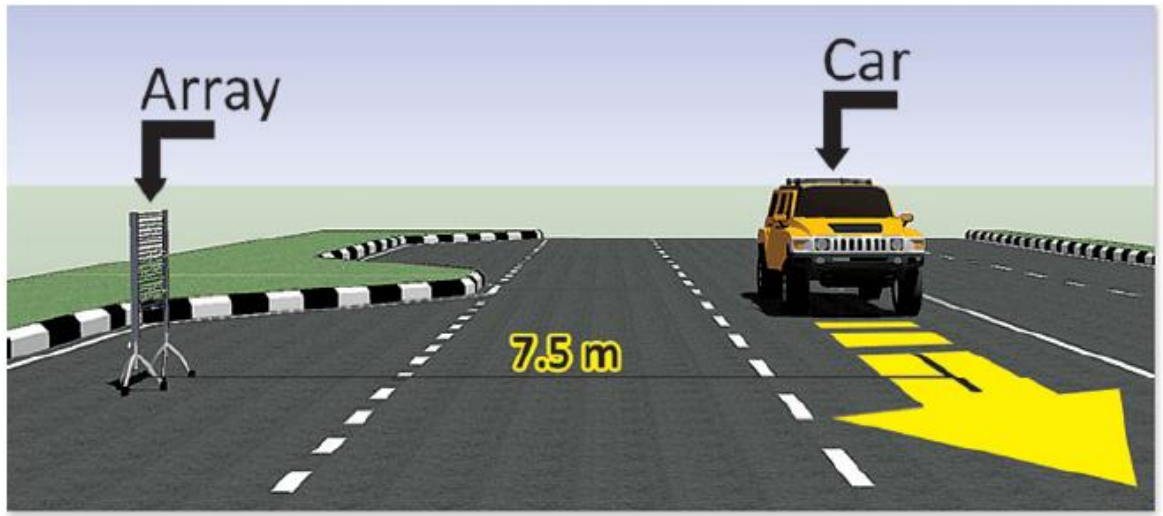


Figure 17: Pass-by Noise Test, beamforming layout. 3D View.

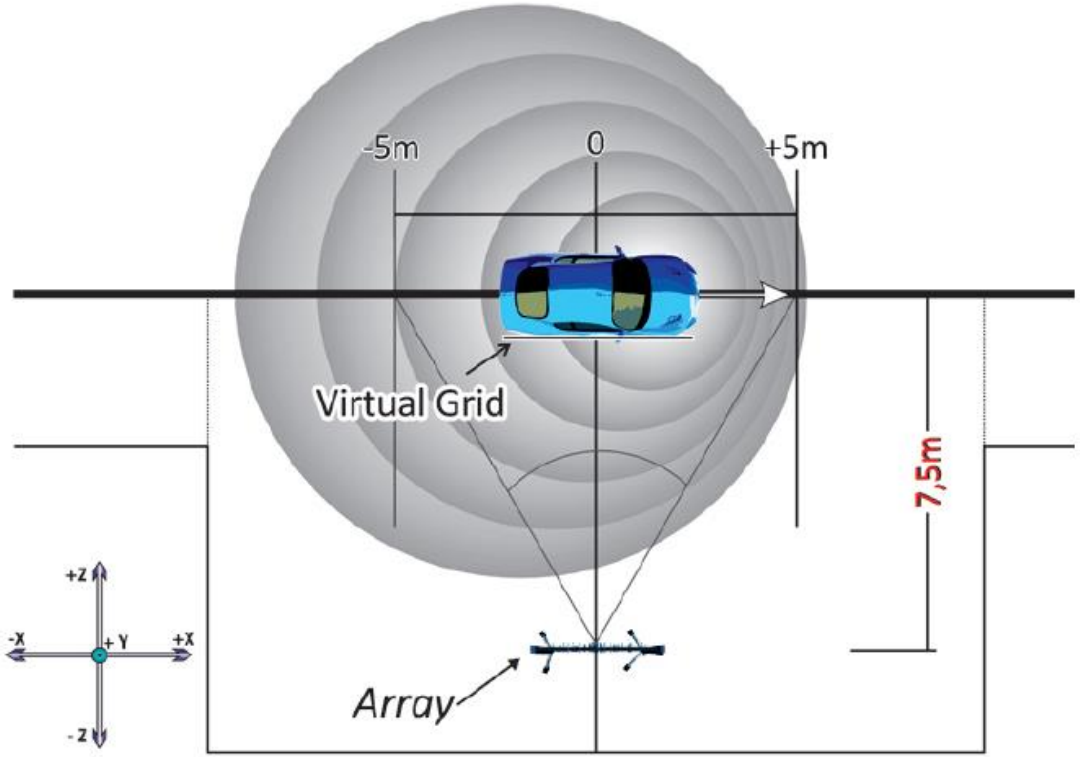


Figure 16: Pass-by Noise Test, beamforming layout.

Beanforming – przykład dla źródeł w ruchu

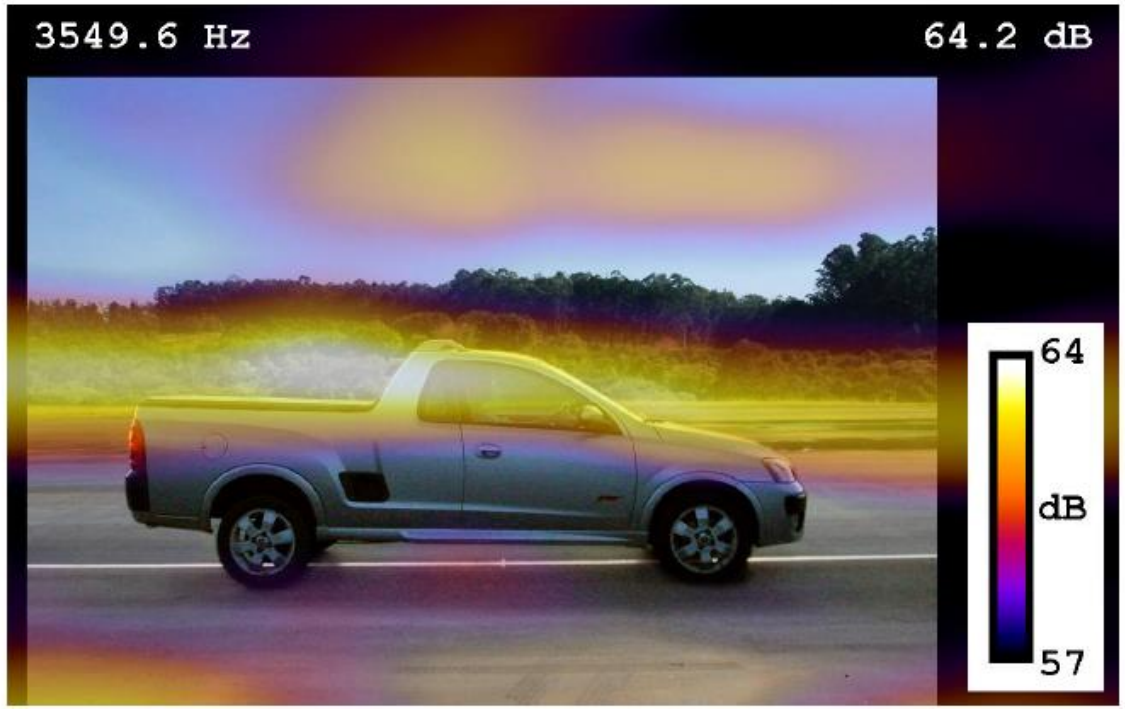


Figure 5: Beamforming data without de-dopplerization.

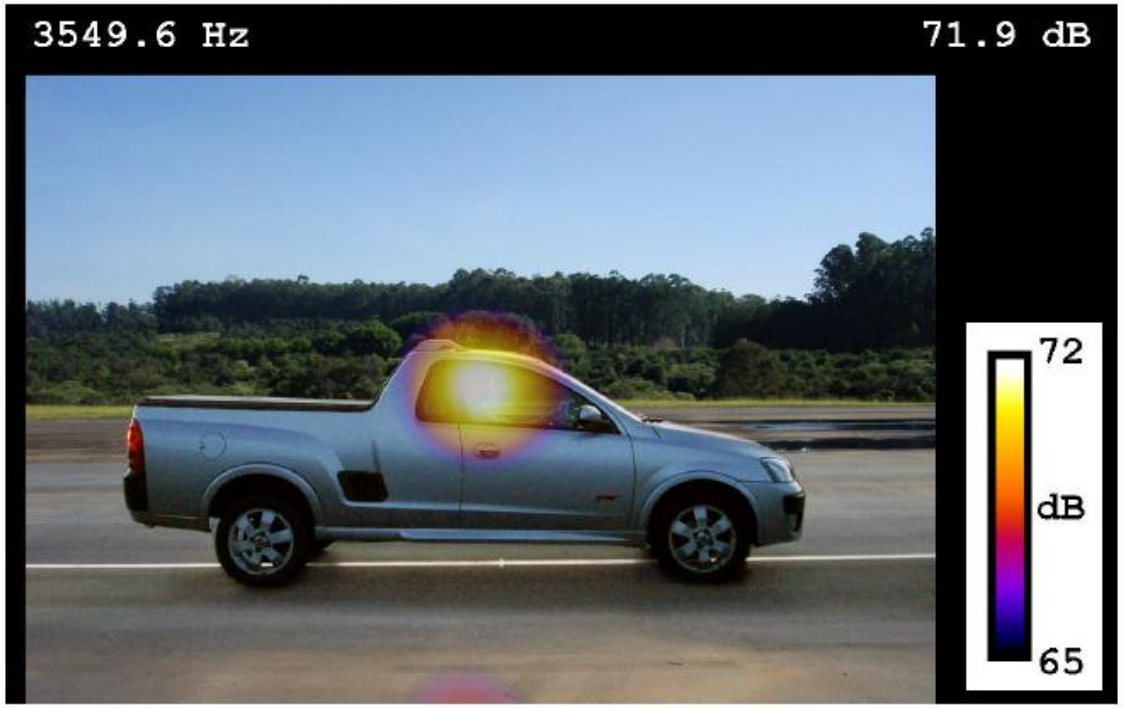
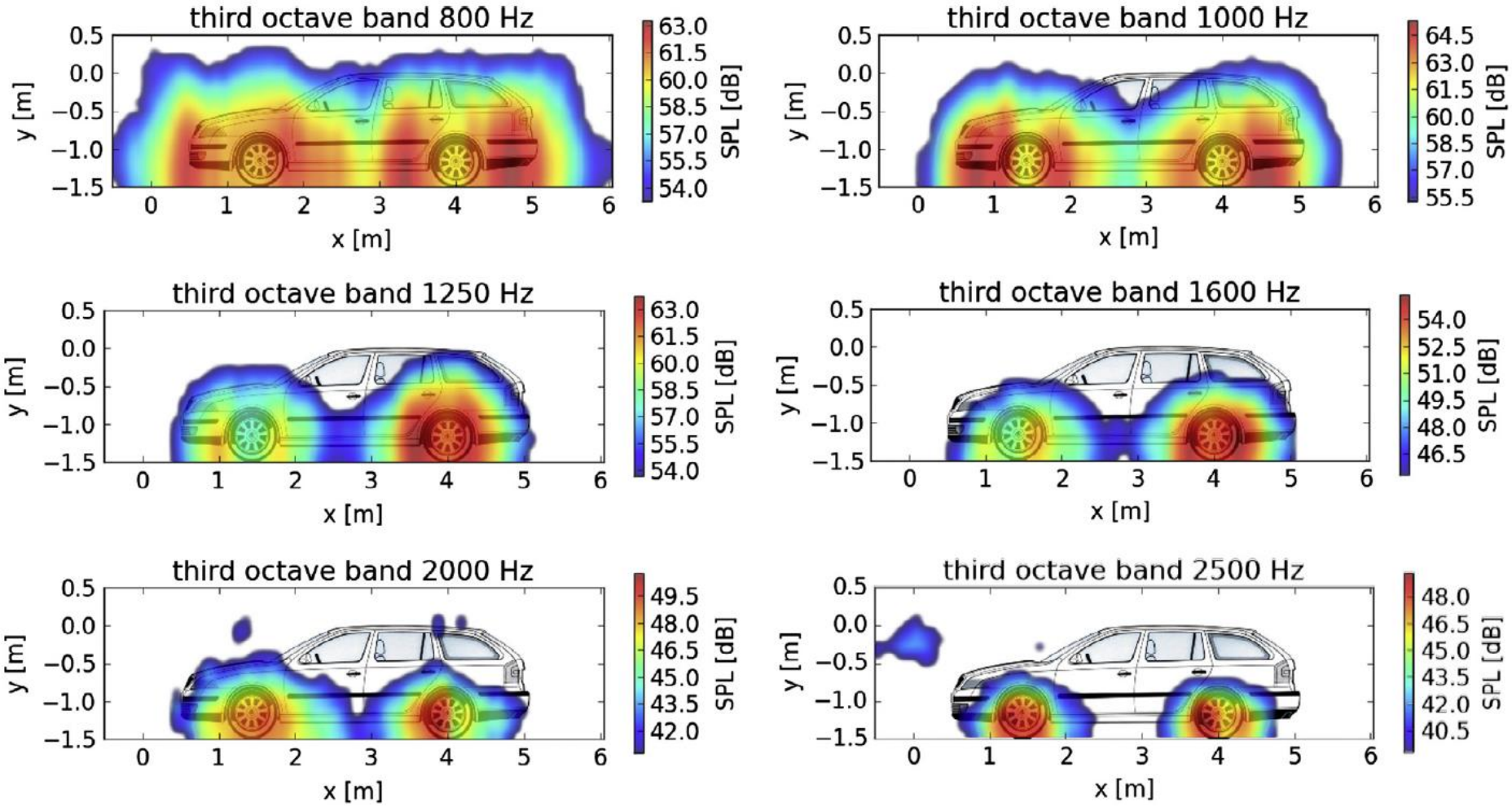


Figure 6: Beamforming data with de-dopplerization.

Beanforming – przykład dla źródeł w ruchu

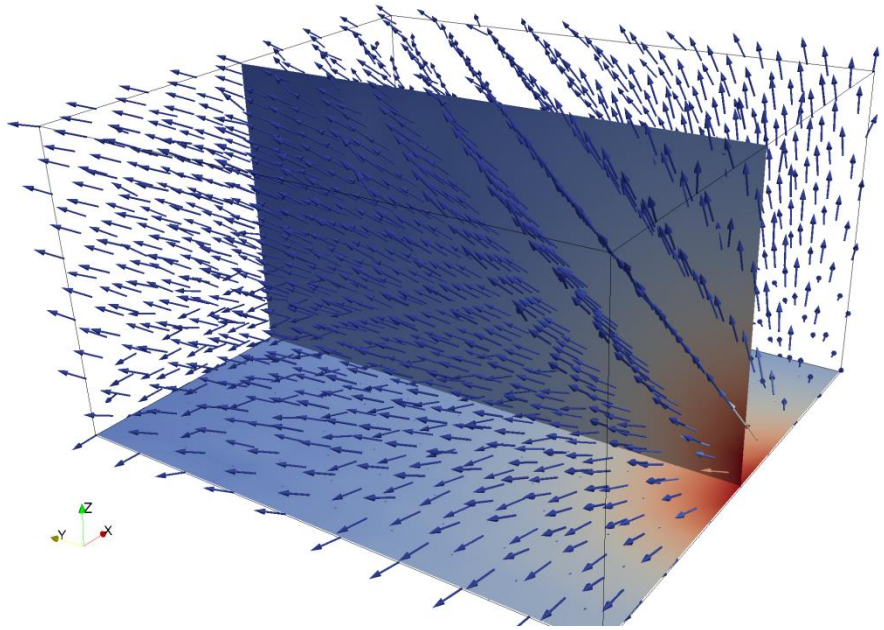
J.A. Ballesteros et al./Applied Acoustics 93 (2015) 106–119



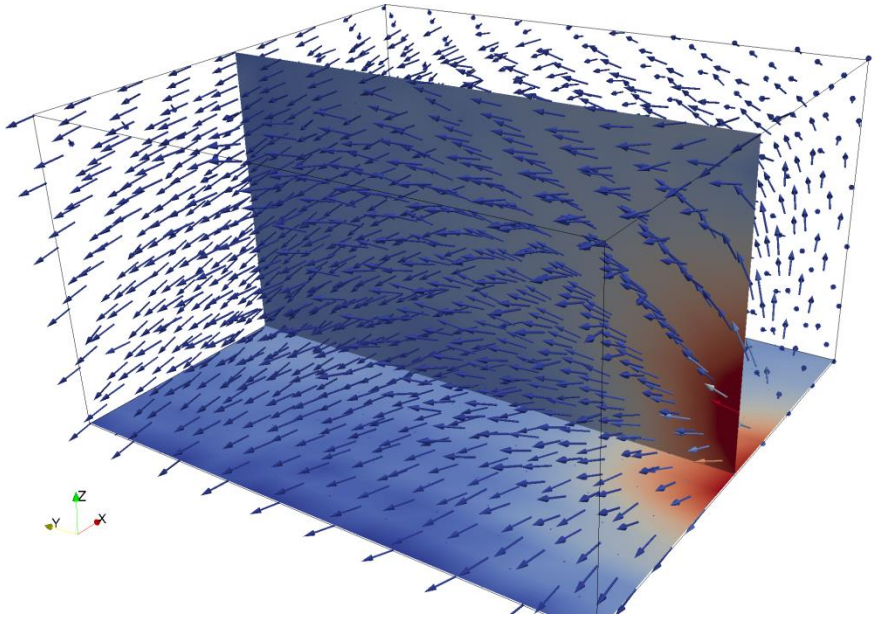
Obrazowanie kierunkowości promieniowania na przykładzie piszczałki organowej




Obrazowanie kierunkowości promieniowania na przykładzie piszczałki organowej

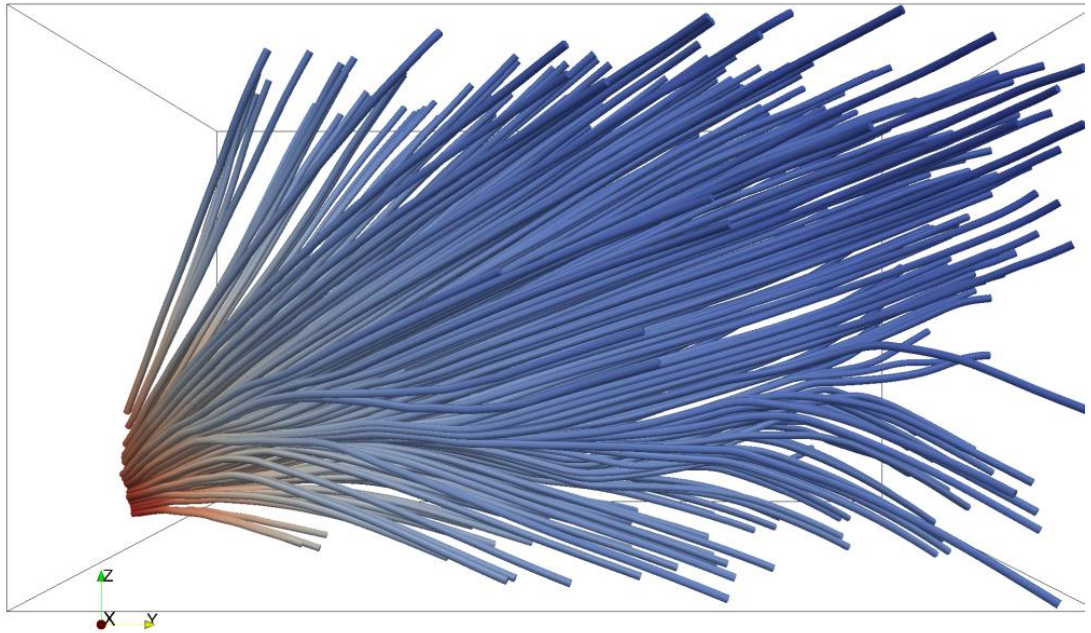


piszczałka drewniana

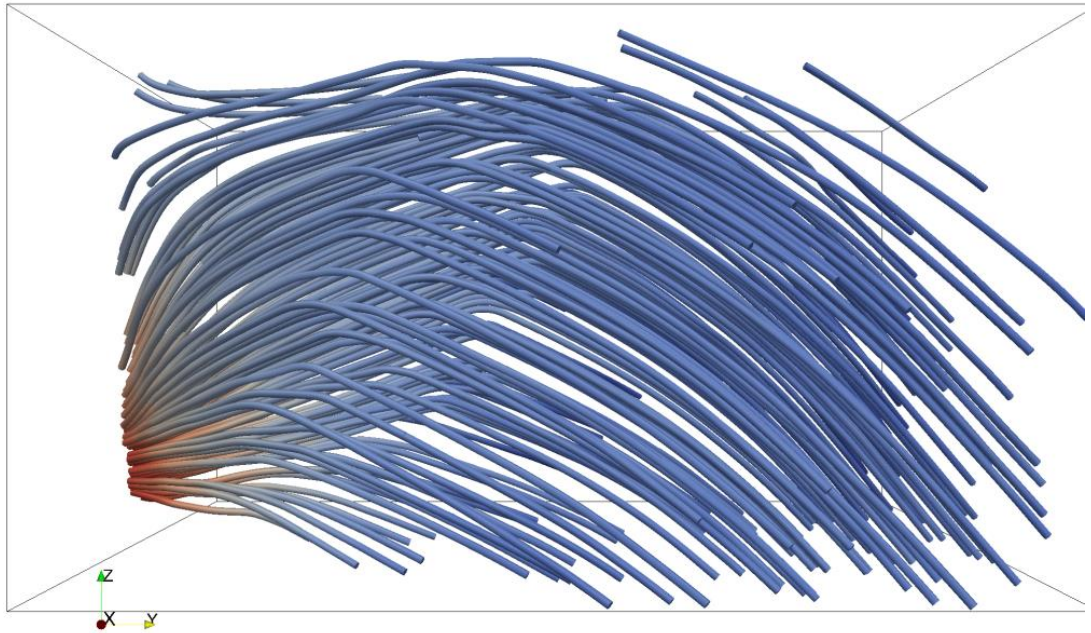


piszczałka metalowa

Obrazowanie kierunkowości promieniowania na przykładzie piszczałki organowej

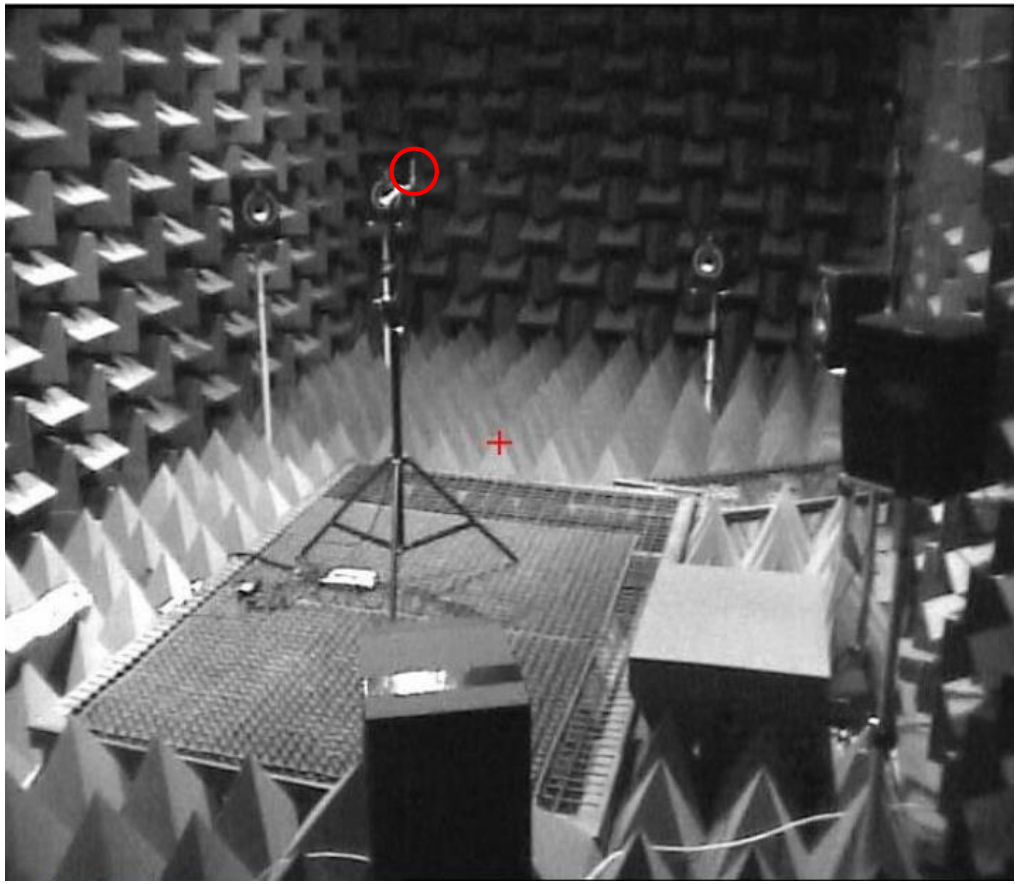


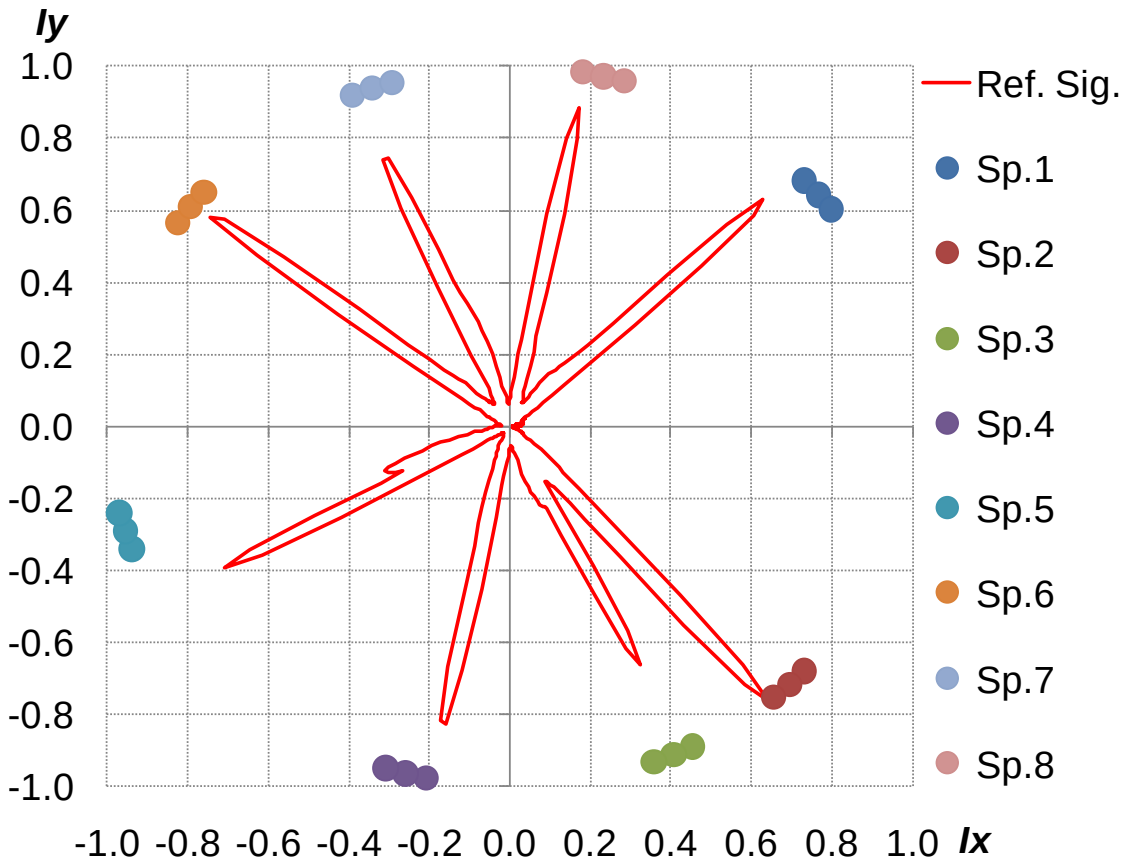
piszczałka drewniana



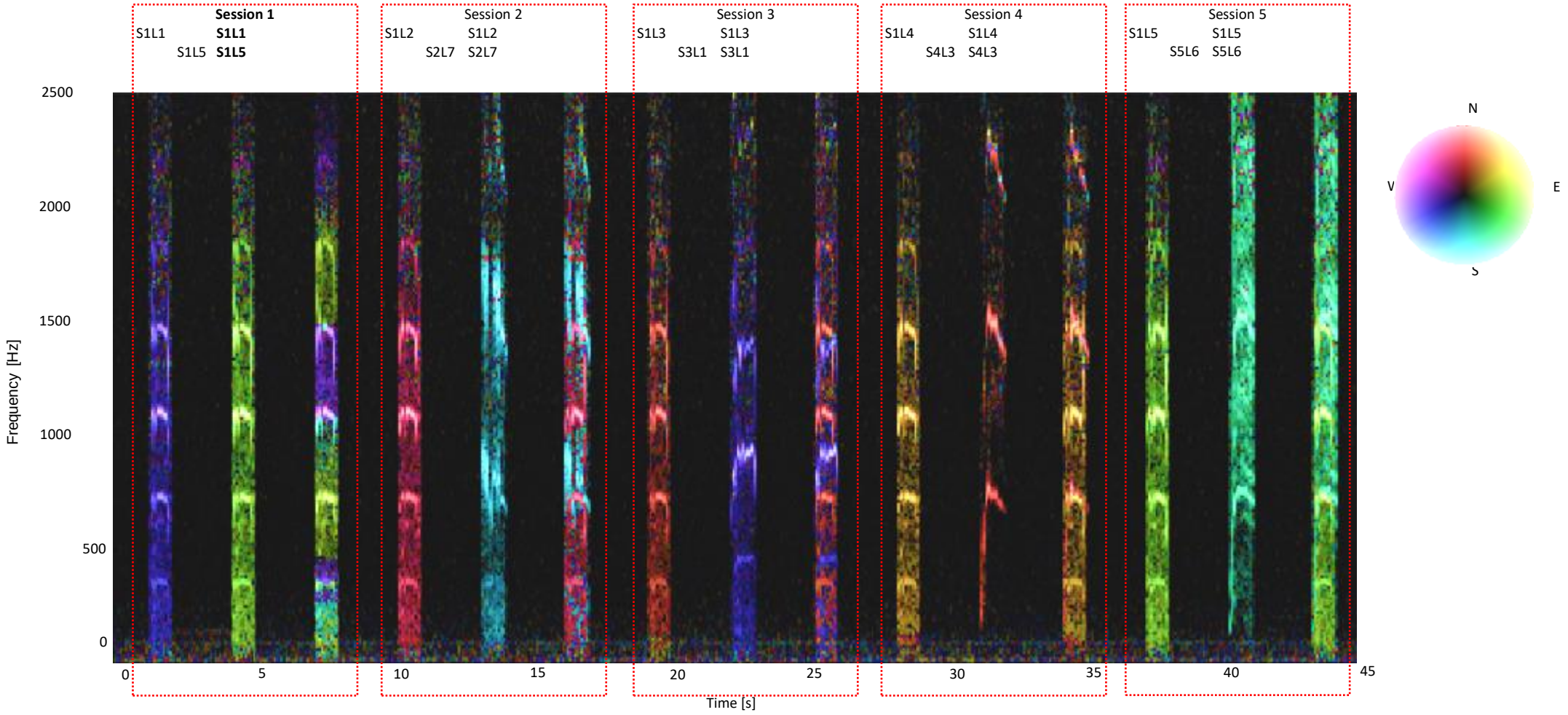
piszczałka metalowa

Obrazowanie kierunku dobiegania dźwięku





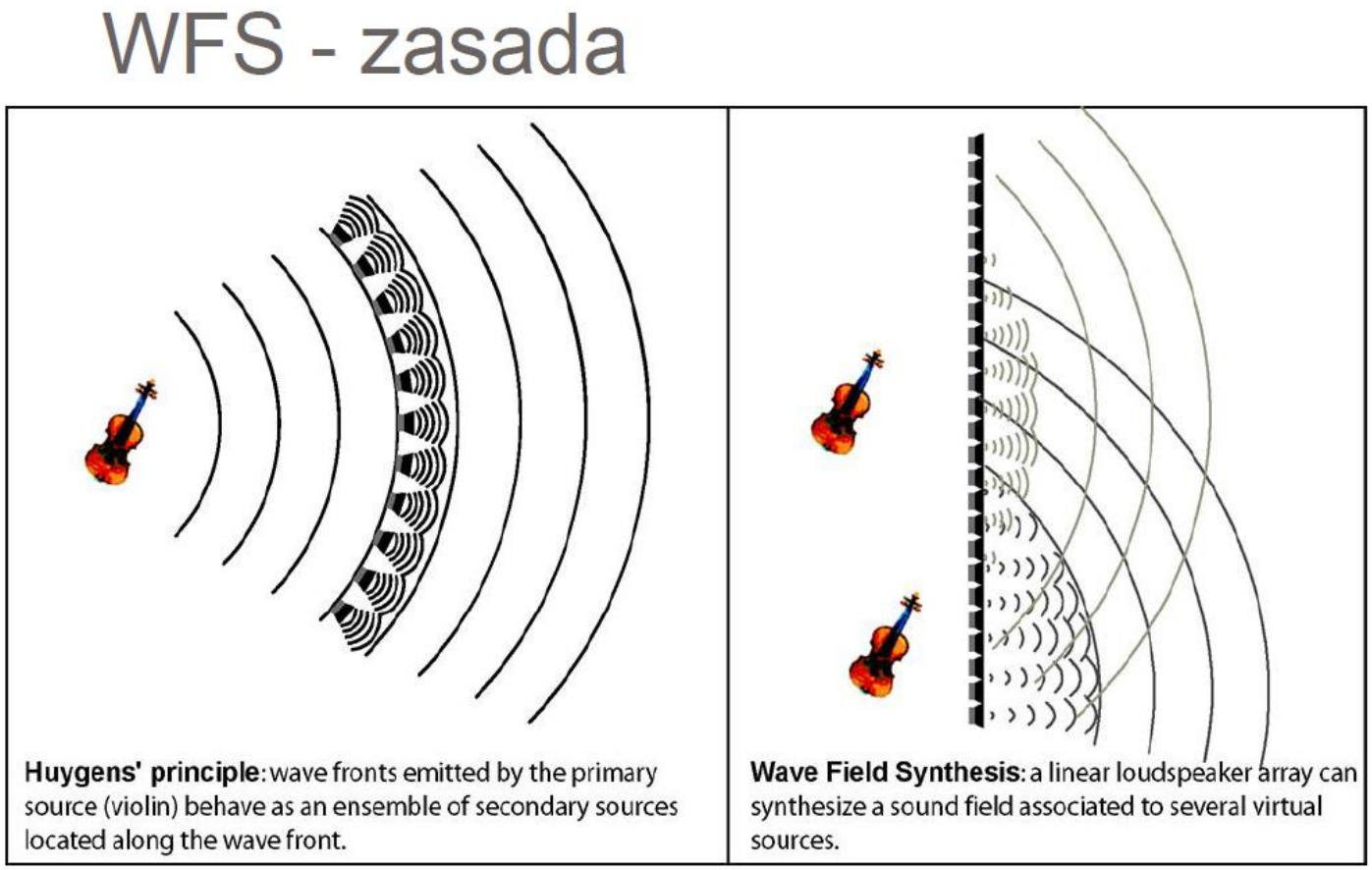
Obrazowanie kierunku dobiegania dźwięku



Beamforming po stronie nadawczej

Wave Field Synthesis

- Wave Field Synthesis (WFS) – technika odwzorowania przestrzennego rozkładu pola akustycznego
- duża liczba głośników (często więcej niż 100)
- zasada Huygensa

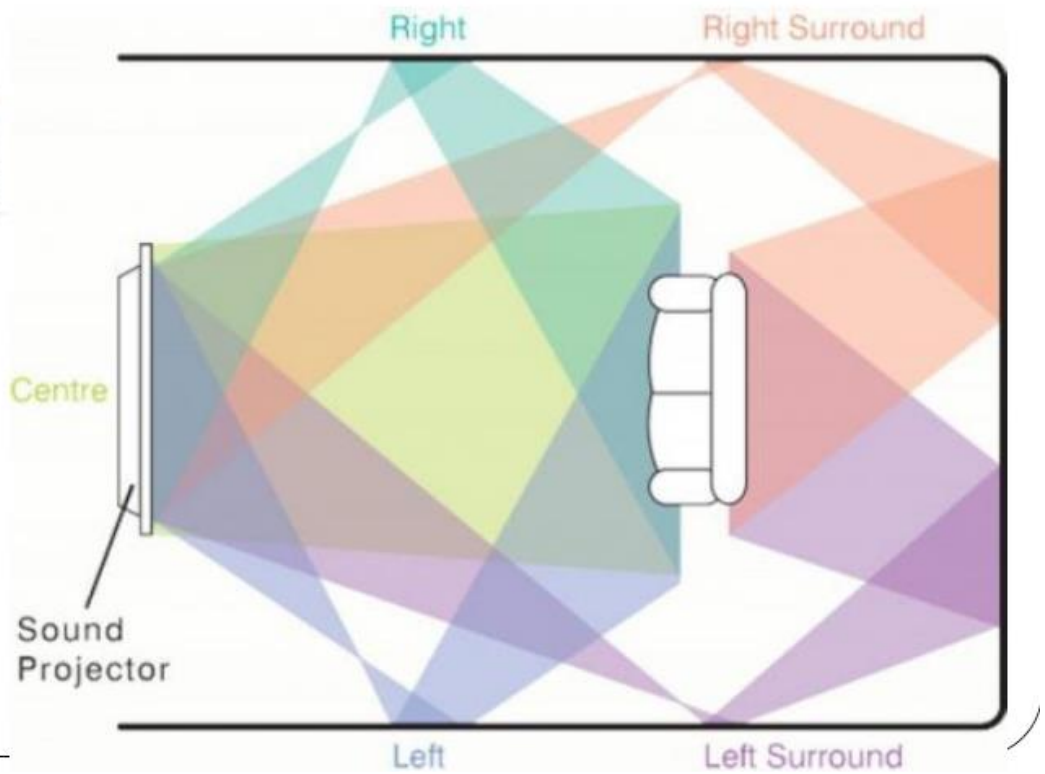


Źródło: recherche.ircam.fr

Projektory dźwięku

- Matryce ok. 40 głośników stosowane w urządzeniach kina domowego – wysoka kierunkowość.

YAMAHA SoundBar



Źródła

- http://www.dydaktyka.cba.pl/InWave.htm
- https://pl.wikipedia.org/wiki/Kszta%C5%82towanie_wi%C4%85zki
- http://www.witczak.imsi.pl/images/stories/04_%20Podstawowe_wielkosci%20_akustyczne.pdf
- http://osilek.mimuw.edu.pl/index.php?title=TTS_Modu%C5%82_9
- https://www.logistyka.net.pl/bank-wiedzy/pozostale-zagadnienia/item/download/76266_7fc5c2bccd76ff456591b12ac61299a5
- https://www.thepodcasthost.com/equipment/microphone-polar-patterns/
- https://www.acoustic-camera.com/en/support/downloads/publications.html
- https://www.ntlmk.com/M.Kirpluk%20-%20Podstawy%20akustyki%20-%202012-11.pdf
- https://www.microflown-avisa.com/technology
- https://charleslabs.fr/en/project-Acoustic+beamsteering+with+a+speaker+array

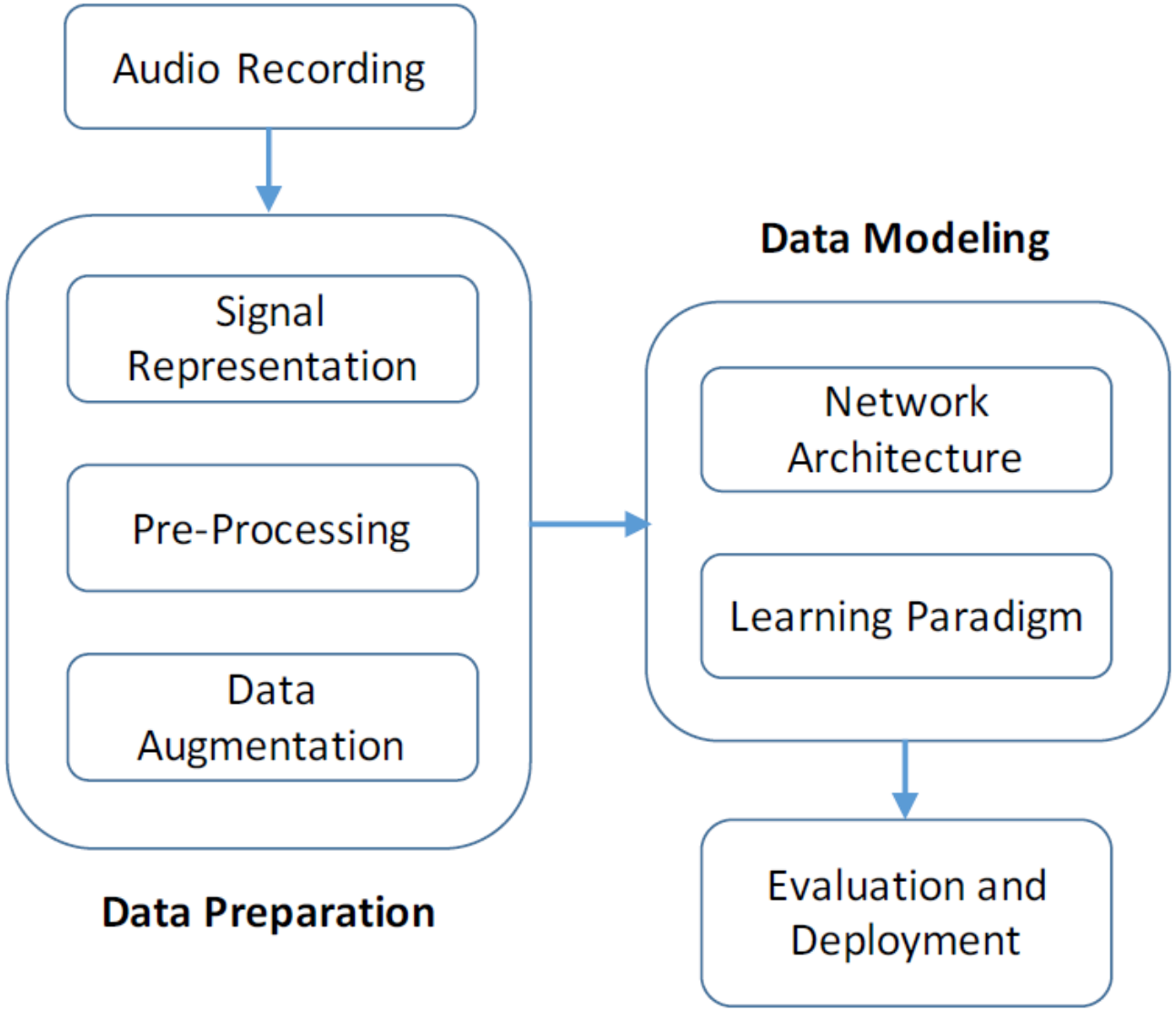
Rozpoznawanie sygnałów fonicznych

Wprowadzenie

- Zasadniczo wyróżniamy dwa podejścia do zagadnienia związanego z detekcją i klasyfikacją zdarzeń akustycznych
  - Podejście klasyczne – klasyfikacja oparta na technikach ekstrakcji cech
  - Podejście oparte na zastosowaniu technik głębokiego uczenia maszynowego, bazujące na różnych reprezentacjach sygnału fonicznego (STFT, MFCC, przekształcenie falkowe…)

Wprowadzenie

- Typowy algorytm przetwarzania sygnału na potrzeby klasyfikacji sygnałów fonicznych



ASC vs AED

- Klasyfikacja scen akustycznych (ASC – Acoustic Scene Classification) *rozpoznawanie różnych środowisk akustycznych wewnątrz i na zewnątrz na podstawie zarejestrowanych sygnałów akustycznych, np.: „pokój biurowy” lub „miejsce publiczne”*
- Wykrywanie zdarzeń akustycznych (AED - Acoustic Event Detection) *wykrywanie zdarzeń dźwiękowych, które są tymczasowo obecne w scenie akustycznej. Przykłady takich zdarzeń dźwiękowych obejmują: odgłosy pojazdów, klaksony samochodowe, kroki, pośród innych*
- AED zasadniczo różni się od ASC, ponieważ skupia się na precyzyjnym czasowym wykrywaniu poszczególnych zdarzeń dźwiękowych

ASC – Acoustic Scene Classification

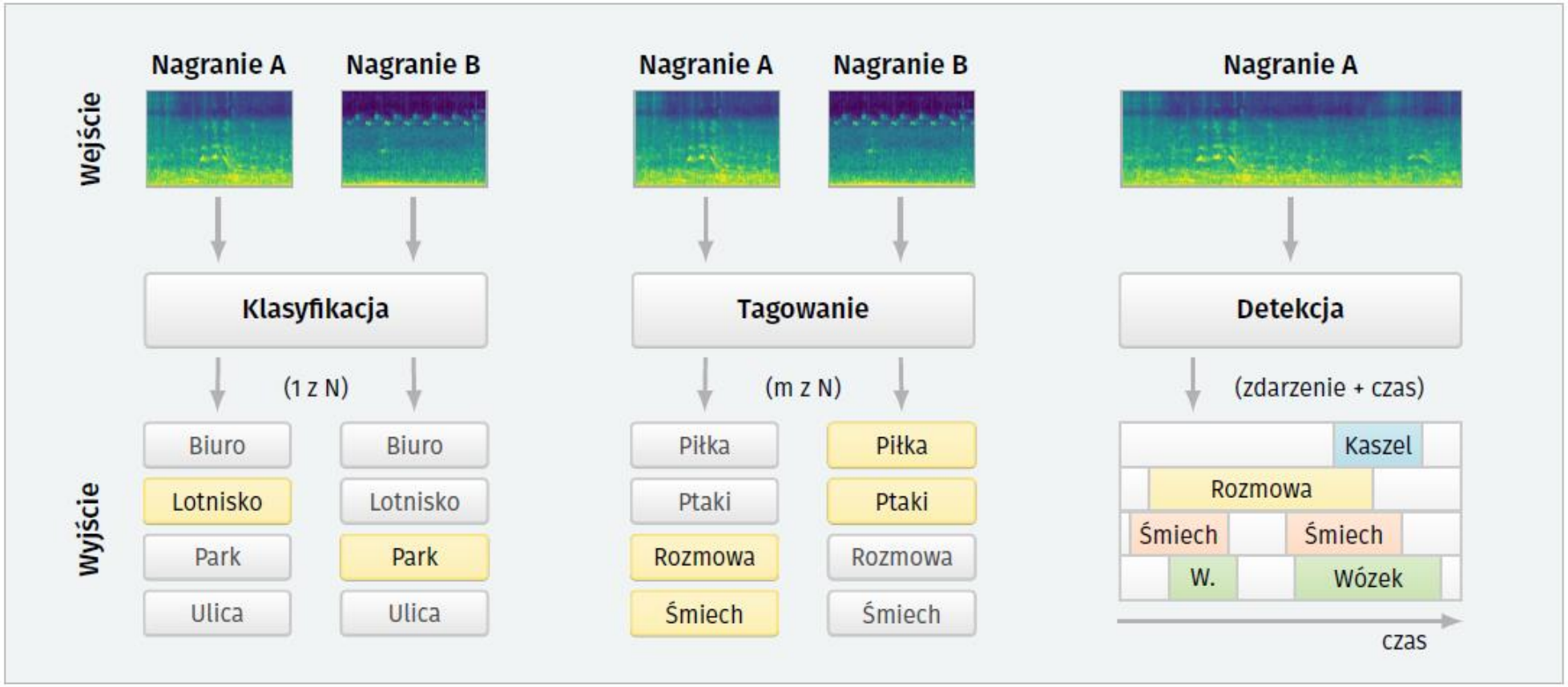
- *Wykazano, że najnowocześniejsze systemy ASC przewyższają ludzi w tym zadaniu [2]. W związku z tym, są stosowane w wielu scenariuszach aplikacji, takich jak kontekstowe urządzenia do noszenia i słuchu, aparaty słuchowe, opieka zdrowotna, nadzór bezpieczeństwa, monitoring dzikiej przyrody w naturalnych siedliskach, inteligentne miasta, IoT i autonomiczna nawigacja.*

dcase.community

- **D**etection and **C**lassification of **A**coustic **S**cenes and **E**vents
  - Serwis internetowy skupiający społeczność osób zaangażowanych w rozwój metod cyfrowego przetwarzania sygnałów dla celów detekcji i klasyfikacji scen i zdarzeń akustycznych przez porównanie różnych podejść przy użyciu wspólnego, publicznie dostępnego zbioru danych.
  - Jedną z motywacji do podjęcia inicjatywy wspólnego rozwiązywania ww. zagadnień jest ograniczenie problemów związanych z trudnością wiarygodnej oceny i porównanie skuteczności różnych metody poprzez udostępnienie obszernego repozytorium danych testowych

https://www.youtube.com/@DCASEWorkshop/videos

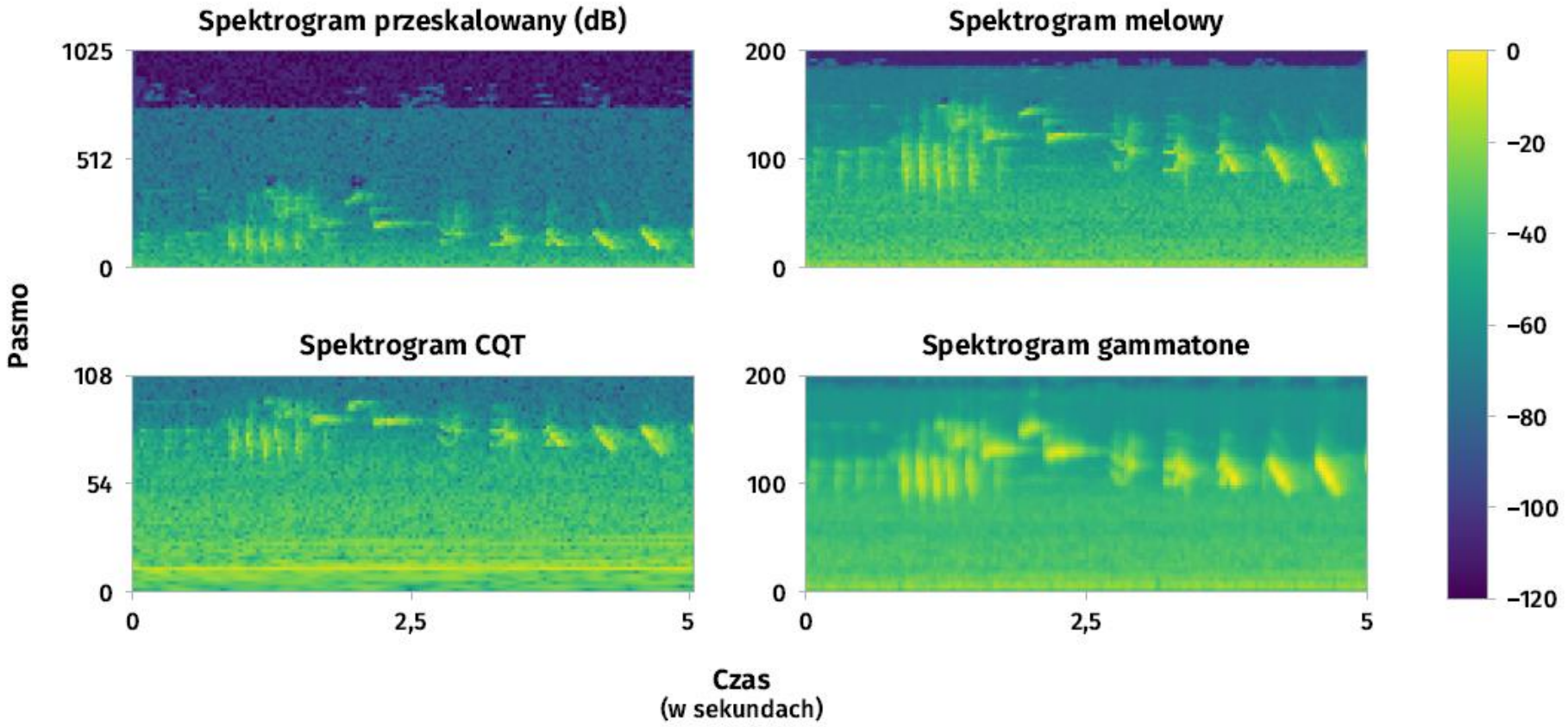
Typy zadań rozpoznawania dźwięku według systematyki Virtanena et al. (2018).



Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

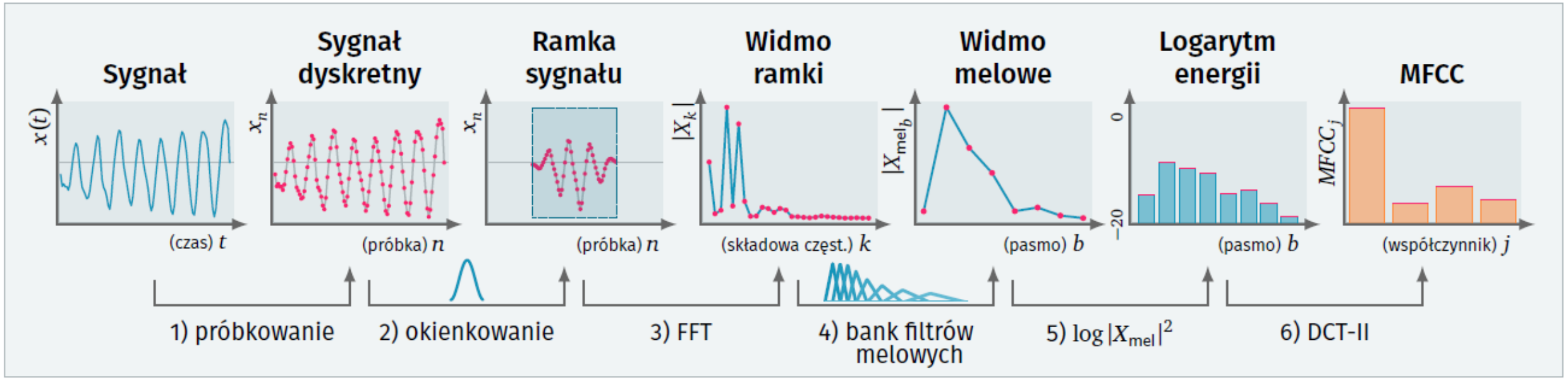
Porównanie reprezentacji nagrania za pomocą różnych wariantów spektrogramów



Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Schemat wyznaczania współczynników mel-cepstralnych (MFCC)



Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Ocena jakości systemu rozpoznawania dźwięku

Porównanie wygenerowanych predykcji z wartościami rzeczywistymi pozwala na wyodrębnienie następujących sytuacji:

- wynik prawdziwie pozytywny (**true positive, TP**) - zarówno model i dane referencyjne wskazują na występowanie danego zdarzenia w analizowanym segmencie,
- wynik fałszywie pozytywny (**false positive, FP**) - model przewiduje wystąpienie zdarzenia dźwiękowego, które nie jest odzwierciedlone w danych referencyjnych,
- wynik fałszywie negatywny (**false negative, FN**) - zdarzenie dźwiękowe, które zostało oznaczone w danych referencyjnych, nie zostało wykryte przez model,
- wynik prawdziwie negatywny (**true negative, TN**) - model i dane referencyjne jednocześnie wskazują na brak zdarzenia dźwiękowego (wyniki te są rzadko wykorzystywane w dalszym obliczaniu miar jakości).

Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Ocena jakości systemu rozpoznawania dźwięku

Jeżeli przez ***TP***, ***FP***, ***FN***, ***TN*** oznaczymy odpowiednio liczby wystąpień wyników prawdziwie pozytywnych, fałszywie pozytywnych, fałszywie negatywnych i prawdziwie negatywnych, to na ich podstawie możliwe jest wyliczenie najczęściej wykorzystywanych miar końcowych w postaci **precyzji** (***precision***) i **czułości** (***recall***):

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

Precyzja opisuje, w jakim stopniu predykcje generowane przez model są trafne, z kolei **czułość** określa, jak duża część zdarzeń dźwiękowych jest rzeczywiście wychwytywana przez system. Miarą, która łączy te dwa aspekty ewaluacji jest ***F-score***

$$F = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Ocena jakości systemu rozpoznawania dźwięku

Innym sposobem pomiaru skuteczności działania systemu rozpoznawania dźwięku jest **stopa błędu** (***error-rate***), która zlicza liczbę różnic między predykcją a danymi referencyjnymi:

- *S* (*substitutions*) - liczba przykładów, dla których system wskazał inną etykietę niż poprawna,
- *D* (*deletions*) - liczba przykładów, dla których system nie wskazał zdarzenia, choć powinien (wyniki fałszywie negatywne po odliczeniu substytucji),
- *I* (*insertions*) - liczba przykładów, dla których system wskazał zdarzenie, chociaż żadne w rzeczywistości nie wystąpiło (wyniki fałszywie pozytywne po odliczeniu substytucji),
- *N* - ogólna liczba aktywnych przykładów.

$$ER = \frac{S + D + I}{N}$$

Konkretne klasyfikatory wygodniej porównuje się jednak nie za pomocą wykresów, ale w postaci skondensowanej do pojedynczych wartości liczbowych. Do tego celu może służyć wyliczanie obszaru pod krzywą ROC (Area Under Curve, AUC) lub określanie **równej stopy błędu** (Equal Error Rate, **EER**), czyli miejsca, w którym udział wyników fałszywie negatywnych (1- *TPR*) równoważy się liczbowo z udziałem wyników fałszywie pozytywnych (FPR).

Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Ocena jakości systemu rozpoznawania dźwięku

Kolejną miarą oceny jakości modeli predykcyjnych jest dokładność (***accuracy***), która wyraża stosunek prawidłowych odpowiedzi do łącznej liczby przykładów:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Główną wadą dokładności jest to, że z jednakową siłą uwzględnia wyniki prawdziwie pozytywne i prawdziwie negatywne. W przypadku zadania detekcji, w którym analizowane zdarzenia mogą występować stosunkowo rzadko, stan ten jest niepożądany – model, który generuje wyłącznie predykcje typu „brak zdarzenia” może cechować się bardzo wysoką dokładnością przy zerowej użyteczności. Jednym ze sposobów korygowania tej nierównowagi jest wprowadzenie odpowiednich mnożników dla obydwu wariantów w ramach wyliczania dokładności zrównoważonej (***balanced accuracy***, *BACC*)

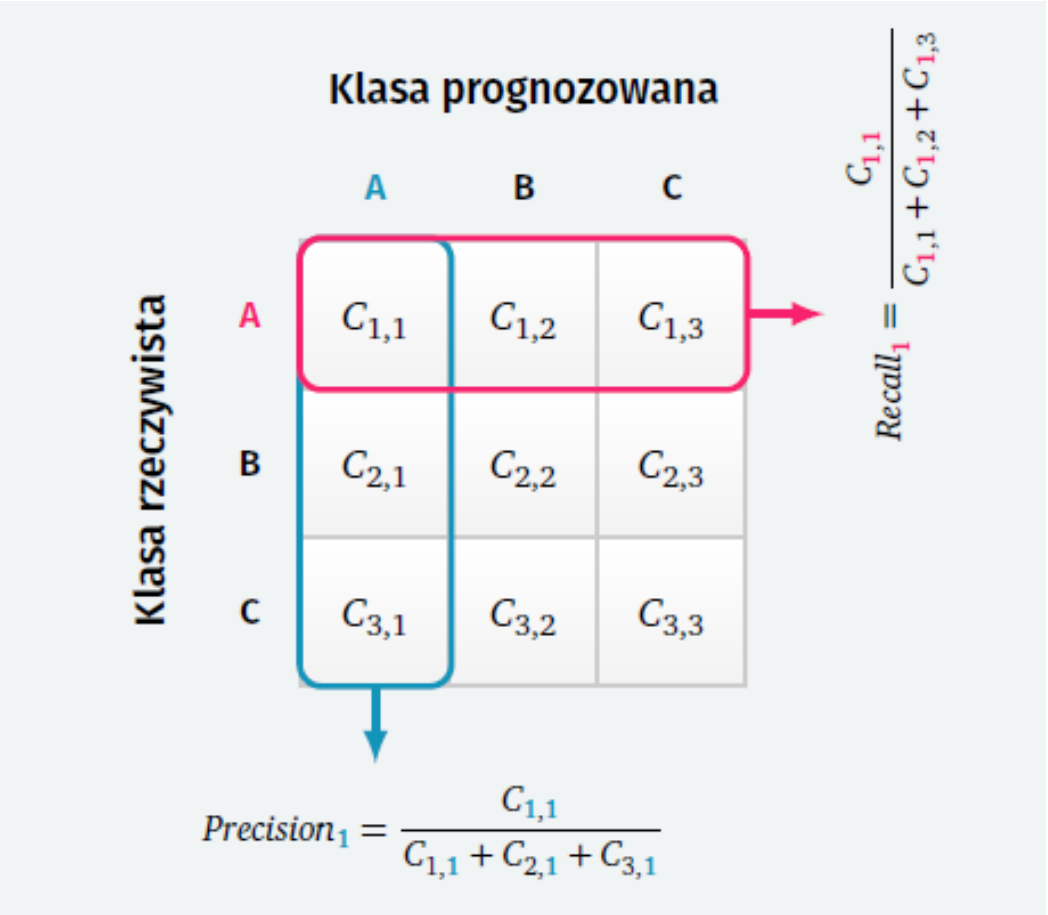
$$BACC = w \cdot \frac{TP}{TP + FN} + (1 - w) \cdot \frac{TN}{TN + FP} = w \cdot TPR + (1 - w) \cdot (1 - FPR)$$

Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Ocena jakości systemu rozpoznawania dźwięku

- **Macierz pomyłek**
  dla klasyfikatora operującego
  na zbiorze 3 klas danych

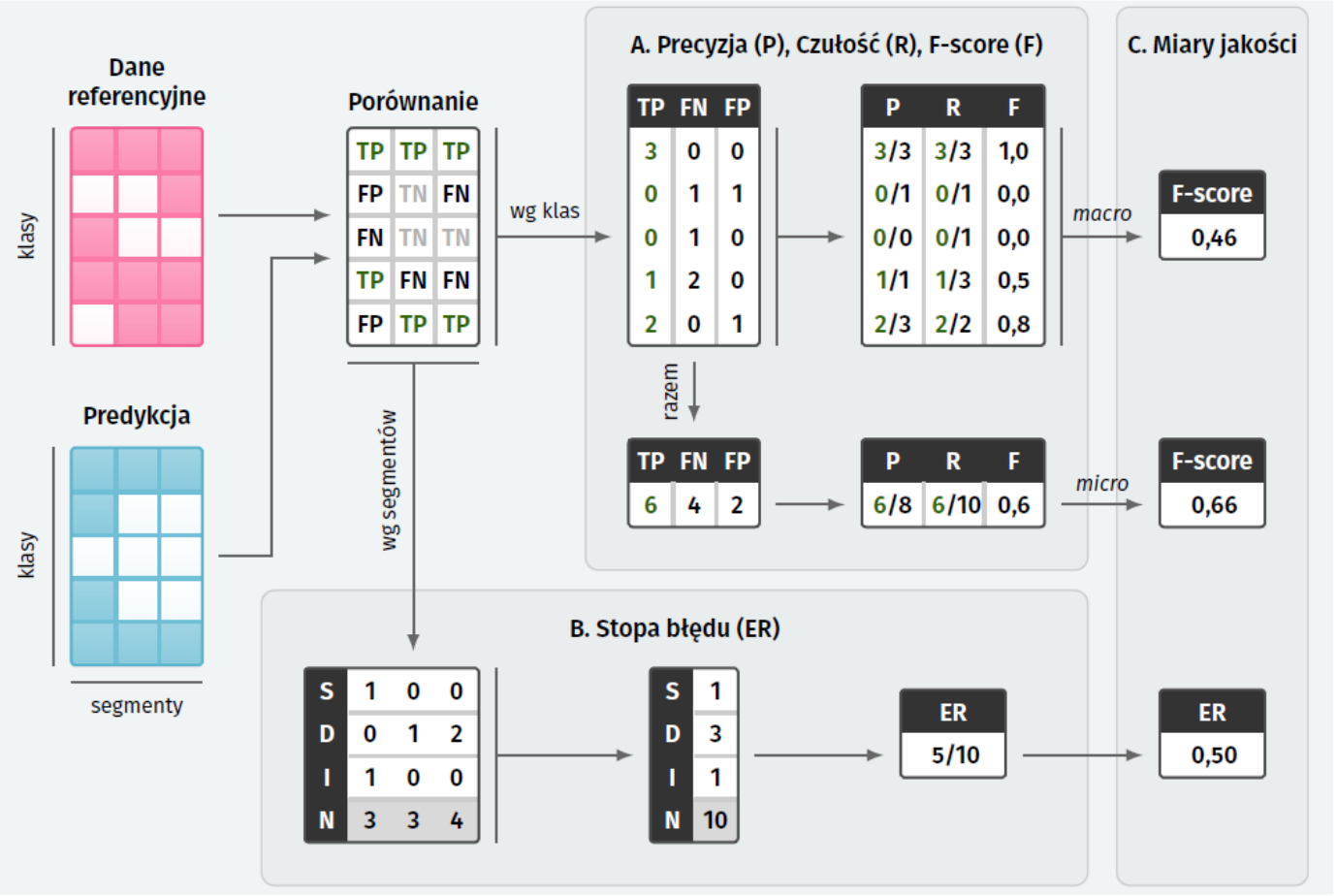


Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Ocena jakości systemu rozpoznawania dźwięku

- Wyznaczanie miar jakości na podstawie analizy segmentów (*segment-based metrics*)

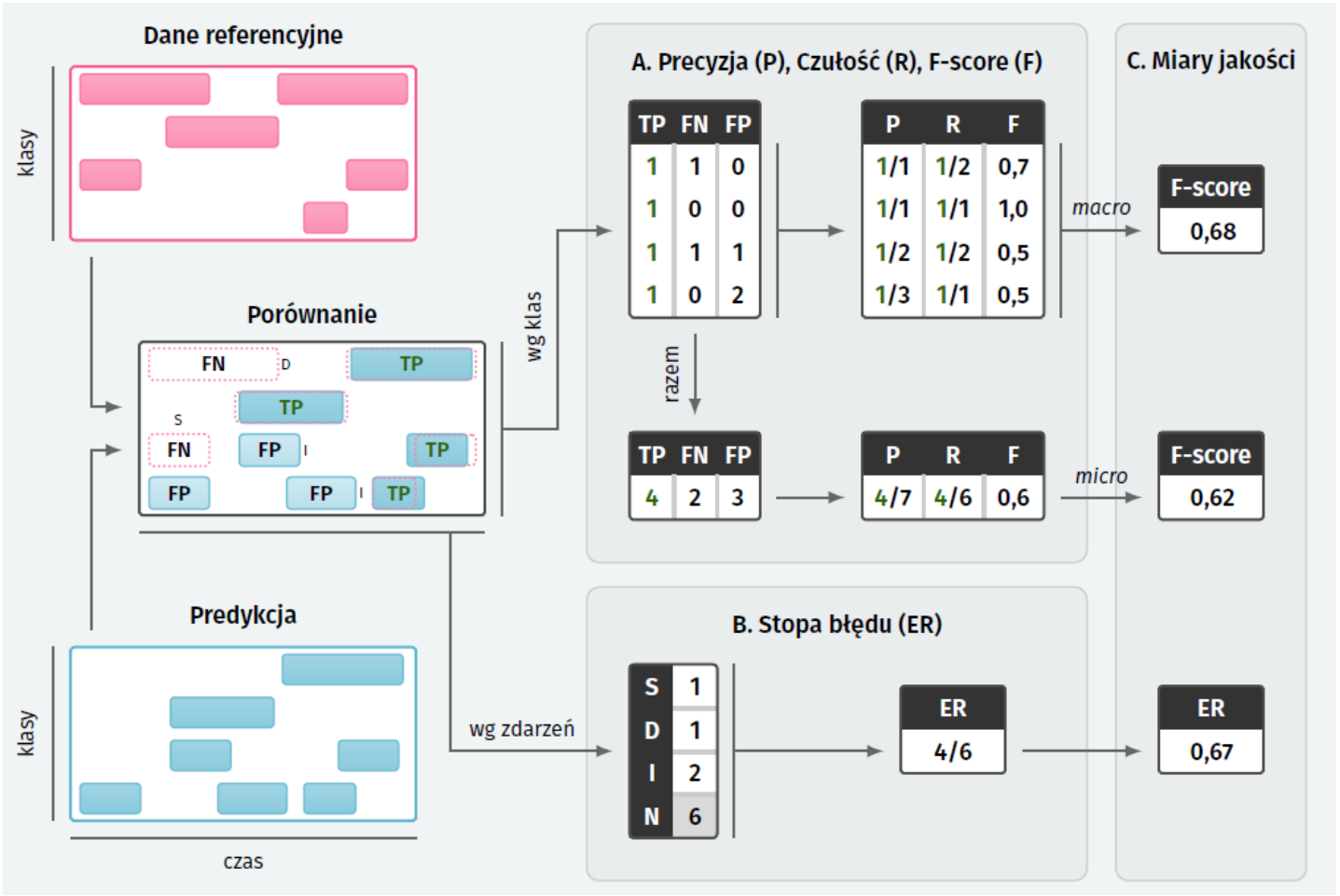


Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Ocena jakości systemu rozpoznawania dźwięku

- Wyznaczanie miar jakości na podstawie poszczególnych zdarzeń (*event-based metrics*)

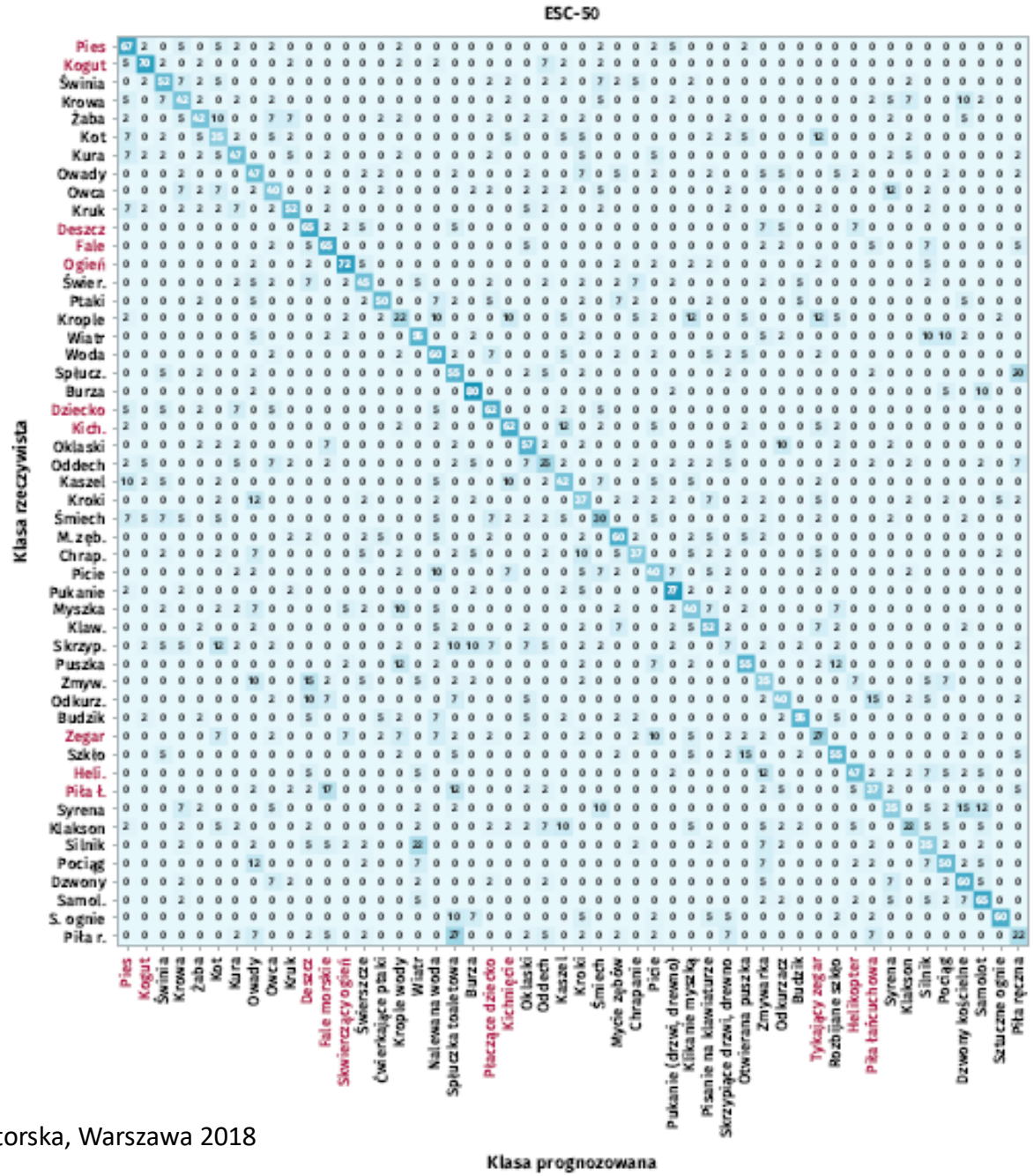


Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Przykład macierzy pomyłek dla wielu klas

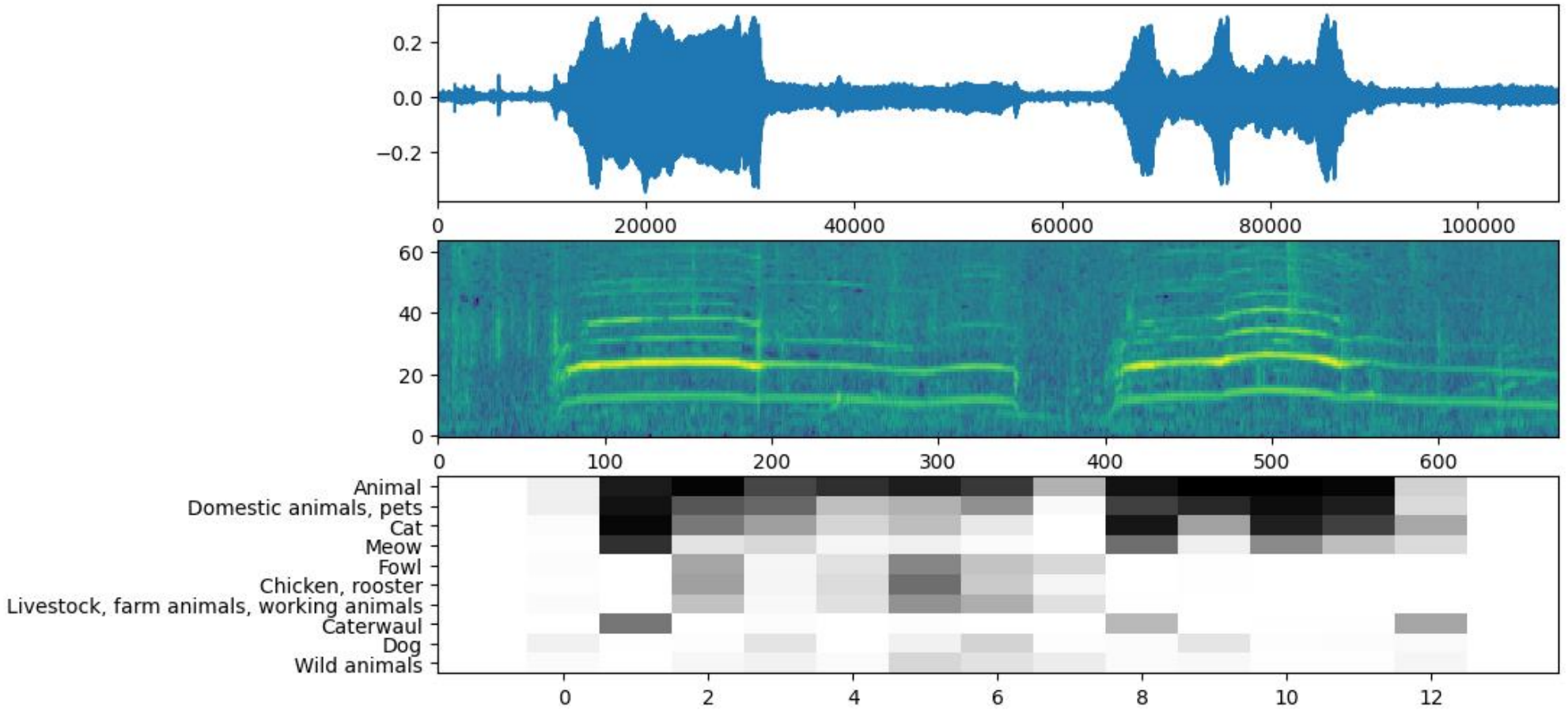
- Macierz pomyłek klasyfikacji zbioru *ESC-50* za pomocą klasyfikatora SVM w wariancie „*C+*”.
- Klasy należące do podzbioru *ESC-10* zaznaczone zostały kolorem ciemnoróżowym.
- Komórki przedstawiają zaokrąglone wartości procentowe.



Karol Jerzy Piczak, Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych, Rozprawa Doktorska, Warszawa 2018

http://repo.pw.edu.pl/info/phd/WUT7e6572509d2e4884b83d981c34379c1c/

Klasyfikacja dźwięku za pomocą YAMNet



https://www.tensorflow.org/hub/tutorials/yamnet

Mapa aktywacji (skupienia)

- Bardzo użytecznym narzędziem do skutecznej interpretacji działania modelu opartego na splotowej sieci neuronowej jest **mapa aktywacji**
- Za jej pomocą można uzyskać informacje, jaka część obrazu została wykorzystana do sformułowania odpowiedzi sieci neuronowej

GradCAM

GradCAM++

https://www.pinecone.io/learn/class-activation-maps/

Źródła

- Abeßer, J. A Review of Deep Learning Based Methods for Acoustic Scene Classification. Appl. Sci. 2020, 10, 2020. https://doi.org/10.3390/app10062020
- http://dcase.community/
- Mesaros, A.; Heittola, T.; Virtanen, T. Assessment of Human and Machine Performance in Acoustic Scene Classification: DCASE 2016 Case Study. In Proceedings of the IEEEWorkshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 15–18 October 2017; pp. 319–323.
- https://www.statystyczny.pl/macierz-bledow-raport-dokladnosc-czulosc-precyzja/
- Łopatka K., Kotus J., Czyżewski A., 2016, Detection, classification and localization of acoustic events in the presence of background noise for acoustic surveillance of hazardous situations, MULTIMEDIA TOOLS AND APPLICATIONS. -Vol. 75, iss. 17, s.1-33, DOI: 10.1007/s11042-015-3105-4
- Mu, W., Yin, B., Huang, X. *et al.* Environmental sound classification using temporal-frequency attention based convolutional neural network. *Sci Rep* **11**, 21552 (2021). https://doi.org/10.1038/s41598-021-01045-4
- https://www.tensorflow.org/hub/tutorials/yamnet
- https://www.pinecone.io/learn/class-activation-maps/
- Kotus, J.; Szwoch, G. Detection of Water on Road Surface with Acoustic Vector Sensor. Sensors 2023, 23, 8878. https://doi.org/10.3390/s23218878
- B. A. Tama et al.: EfficientNet-Based Weighted Ensemble Model for Industrial Machine Malfunction Detection Using Acoustic Signals, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9737110

Dziękuję za uwagę