



AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Zastosowanie techniki kształtowania wiązki w celu zwiększenia skuteczności systemów DSR

*Bartłomiej Chojnacki
Maciej Jagoda*

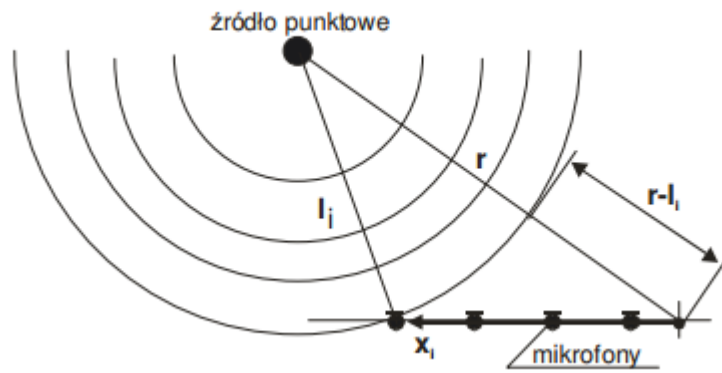
**Wydział Inżynierii Mechanicznej i Robotyki
Katedra Mechaniki i Wibroakustyki**

Kraków, 21.12.2015

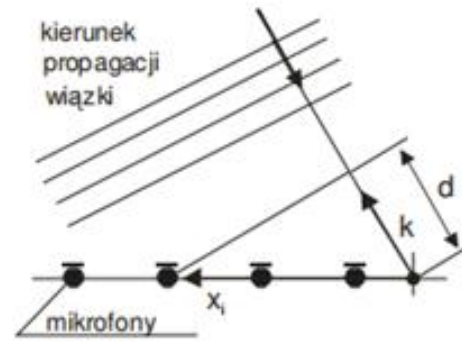
Plan prezentacji

1. Technologia beamformingu i jej zastosowanie w rozpoznawaniu sygnału mowy
2. CHiME-3 - beamforming i filtr Wienera
3. SRLB – sub-band beamforming
4. MCCC - multichannel cross-correlation coefficients
5. Podsumowanie

Beamforming



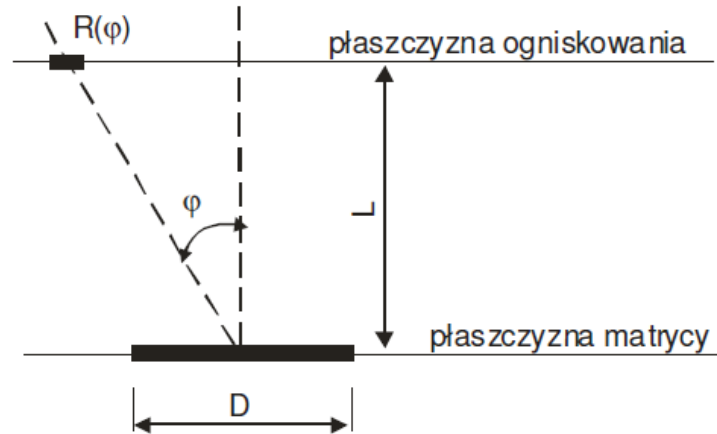
$$\Delta_i = \frac{r - l_i}{c},$$



$$\Delta_i = \frac{d}{c} = \frac{\mathbf{k} \cdot \mathbf{x}_i}{c},$$

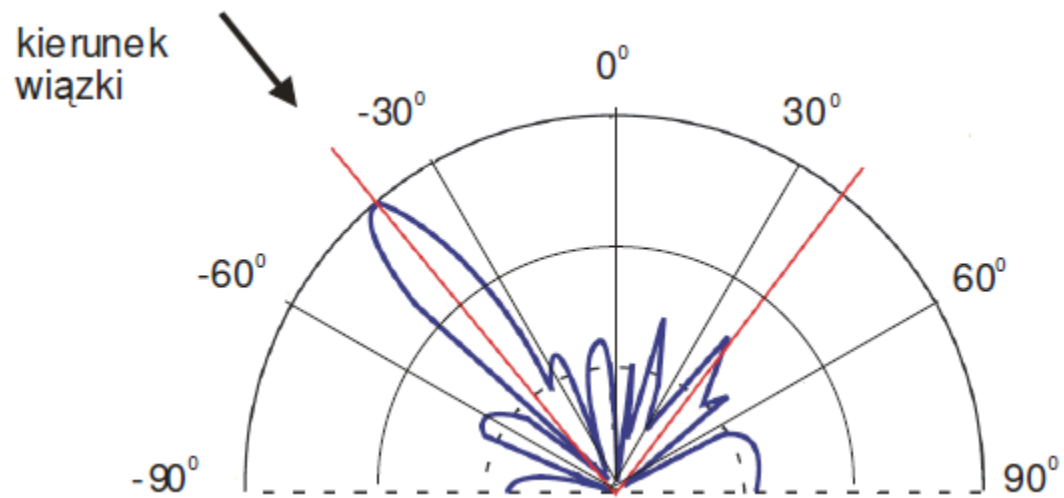
DSR – distant speech recognition

Beamforming

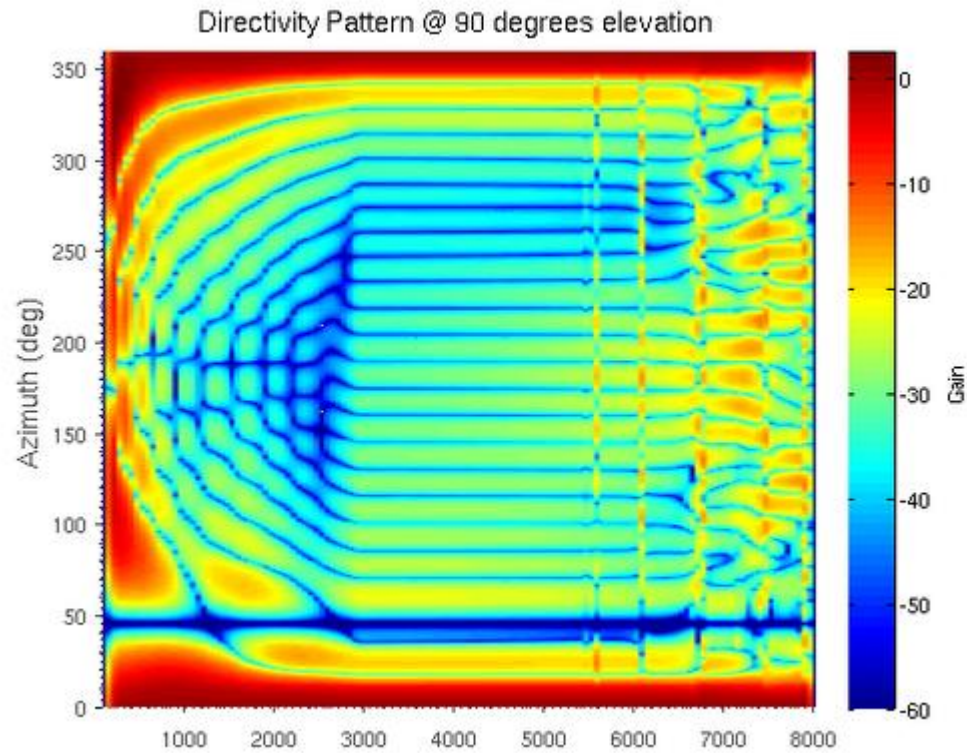


$$R(\varphi) = \frac{a}{\cos^3(\varphi)} \frac{L}{D} \lambda$$

Beamforming

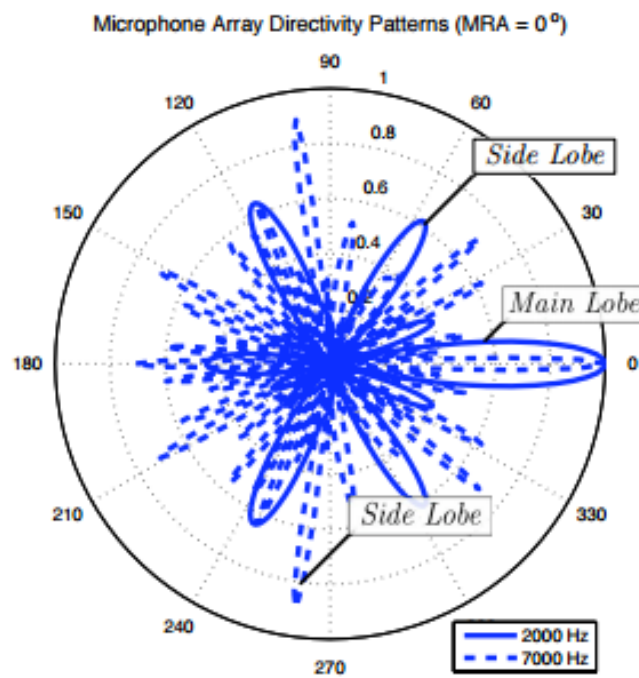


Beamforming

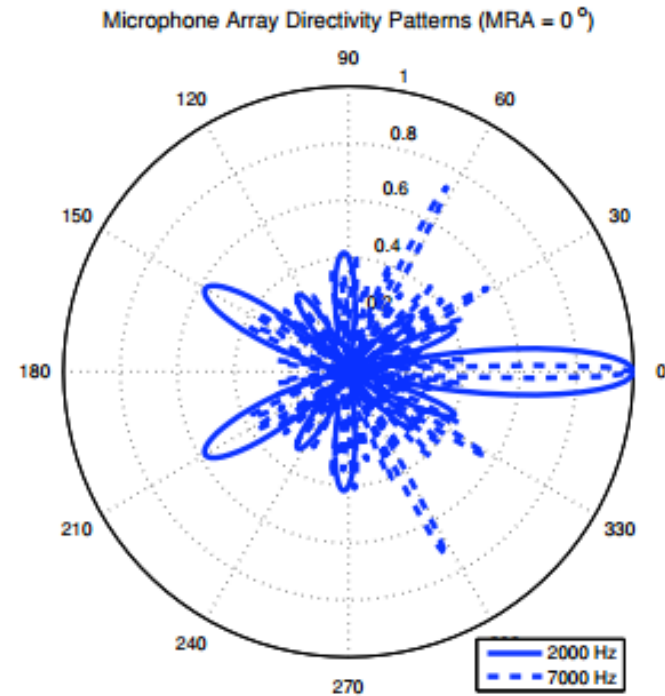


Multiple null synthesis robust least squares frequency invariant beamformer (...eeeeee macarena)

Beamforming

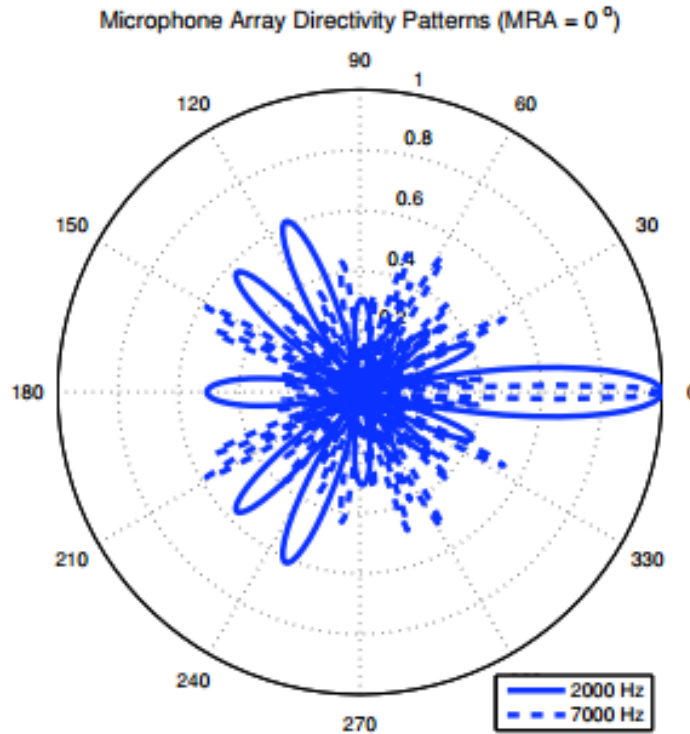


(e) Polar plot of a eight-element array.

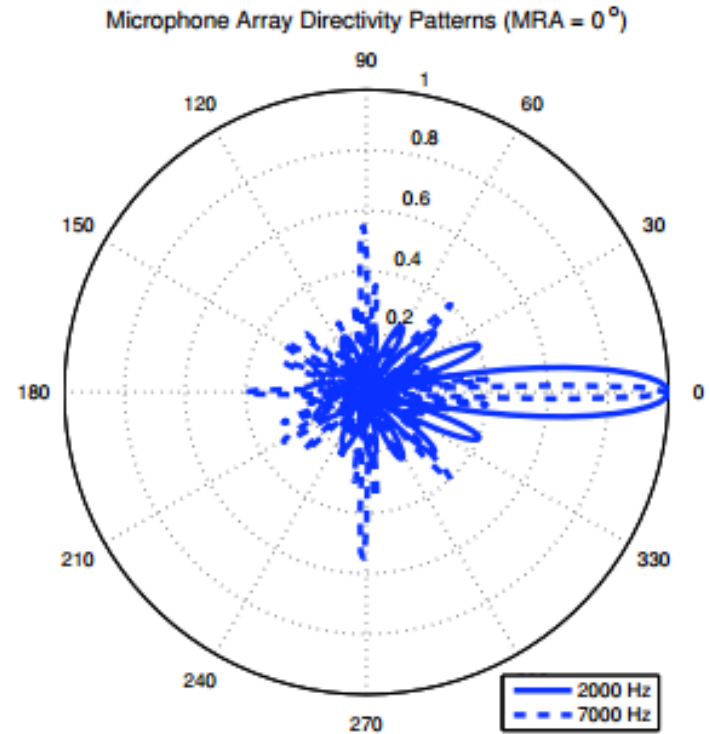


(f) Polar plot of a twelve-element array.

Beamforming

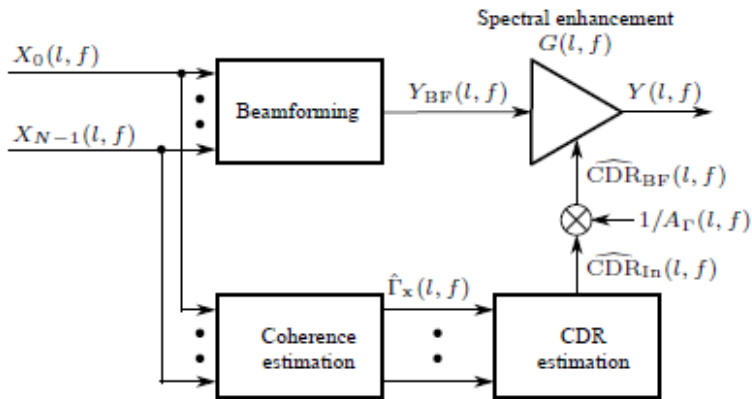


(e) Polar plot of a 16-element array.

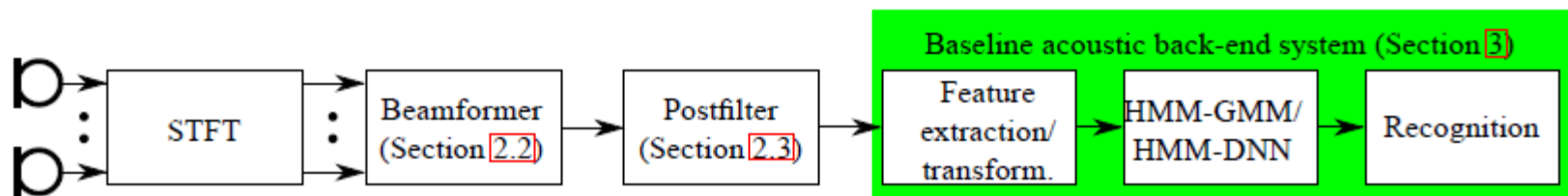


(f) Polar plot of a 24-element array.

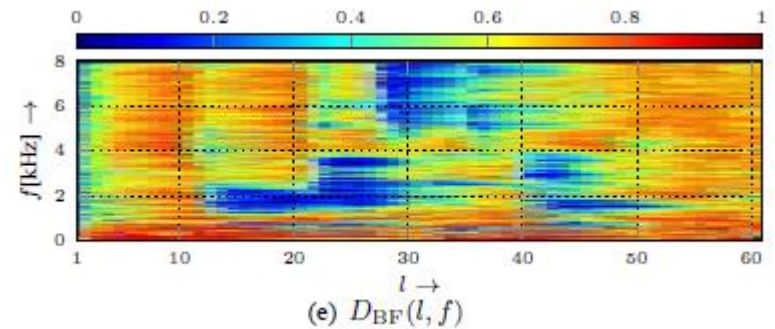
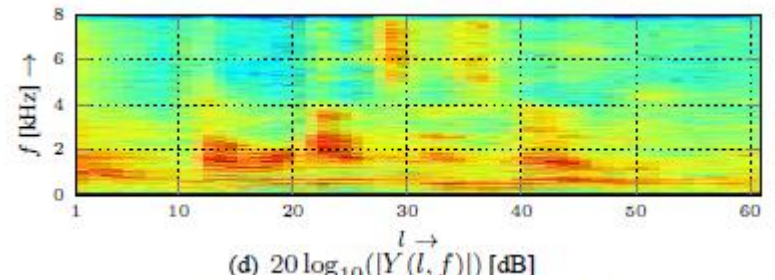
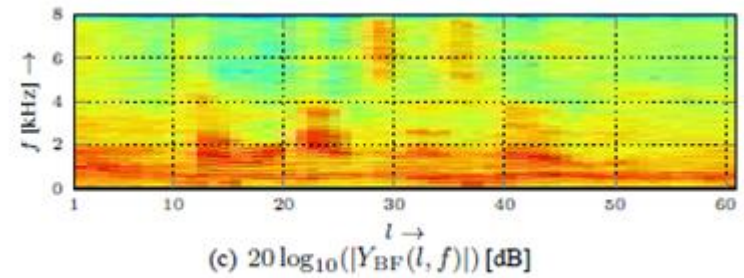
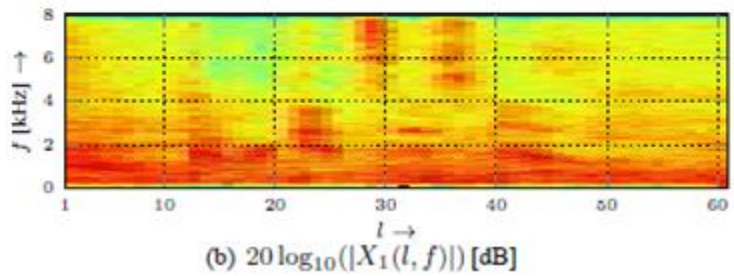
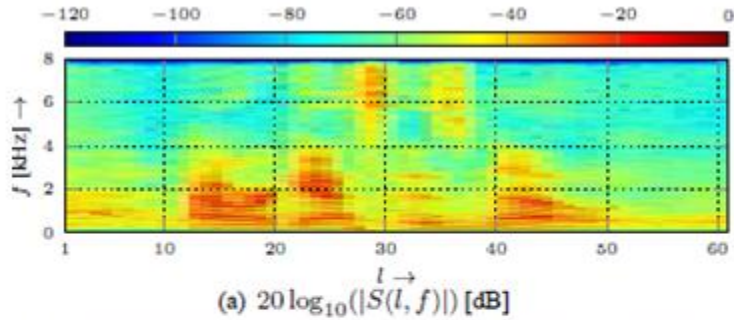
CHiME-3 challenge



„Following the success of the 2011 and 2013 CHiME challenges it gives us great pleasure to pre-announce the 3rd CHiME Speech Separation and Recognition Challenge. The new challenge is designed around the popular Wall Street Journal corpus and features **talkers speaking in challenging noisy environments** recorded using a **6-channel tablet based microphone array.**“



CHiME-3 challenge



Far-Field Continuous Speech Recognition System based on Speaker Localization and Sub-Band Beamforming

Table 1. Speech sub-bands and assigned sub-arrays

Frequency Band	Microphone Index	Microphone Distance
Less than 500	1 - 11	-
500 - 1000	1, 2, 6, 10, 11	20 cm
1000 - 2000	2, 3, 6, 9, 10	10 cm
2000 - 4000	3, 4, 6, 8, 9	5 cm
4000 - 8000	4, 5, 6, 7, 8	2.5 cm

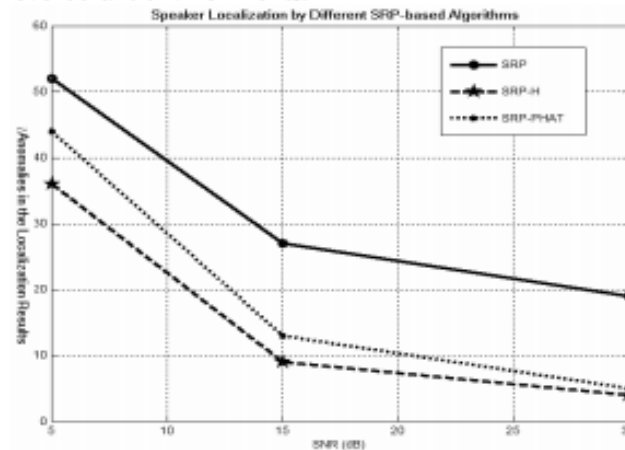
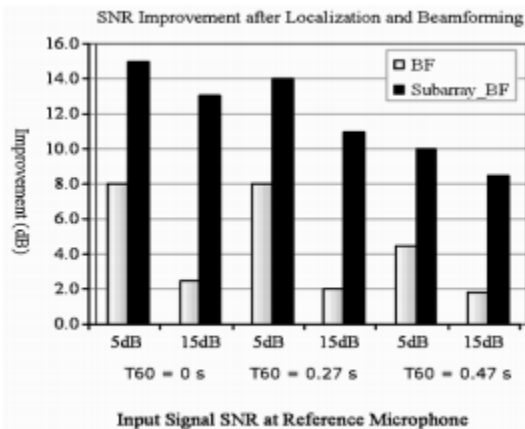
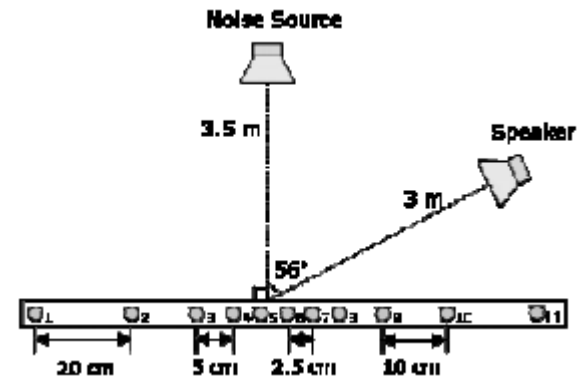


Table 2. Recognition accuracy in the presence of color noise

SNR (dB)	5			15		
	0	0.27	0.47	0	0.27	0.47
T60 (s)						
1. Single (%PAR)	36	22	7	40	25	15
2. BF (%PAR)	9	-2	-18	18	7	-4
3. BF + SL (%PAR)	41	32	11	45	36	27
4. Sub_BF + SL (%PAR)	68	50	27	77	55	30

Table 3. Recognition accuracy in the presence of white noise

SNR (dB)	5			15		
	0	0.27	0.47	0	0.27	0.47
T60 (s)						
1. Single (%PAR)	39	14	2	48	16	6
2. BF + SL (%PAR)	48	20	3	51	24	17
3. Sub_BF + SL (%PAR)	73	34	13	81	39	22

Table 4. Speech recognition in the presence of speech noise and verified speaker localization

LR	1			0.2		
	0	0.27	0.47	0	0.27	0.47
T60 (s)						
1. Single(%PAR)	-4	-9	-16	58	39	33
2. BF (%PAR)	-13	-21	-26	13	12	7
3. BF + SL (%PAR)	2	-2	-6	65	44	37
4. Sub_BF + SL (%PAR)	7	3	-1	75	64	59

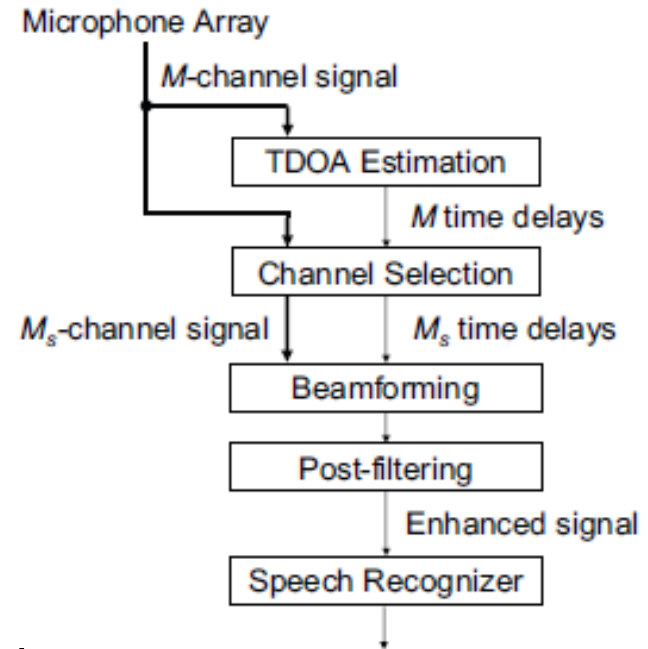
² PAR is calculated by the following equation:

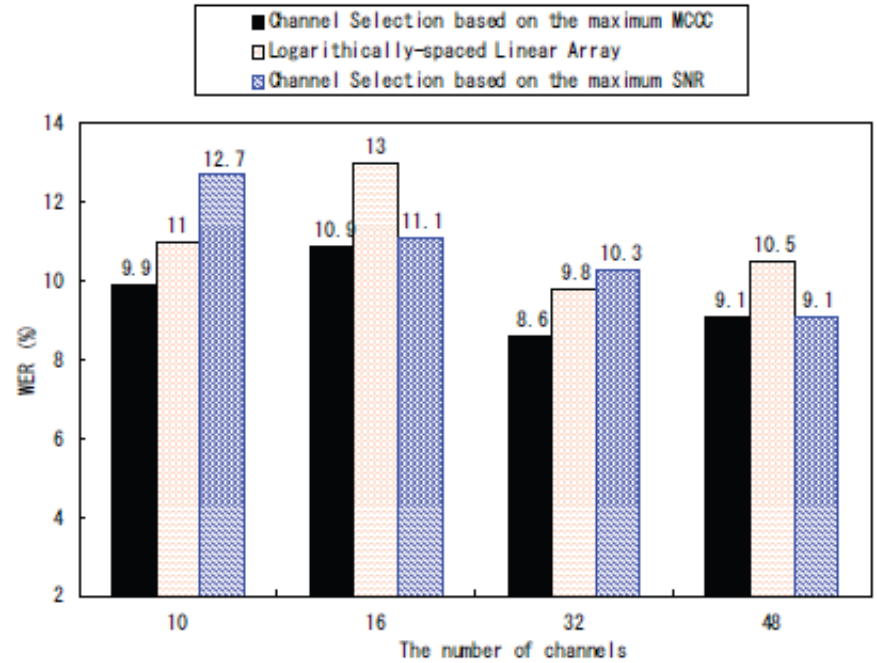
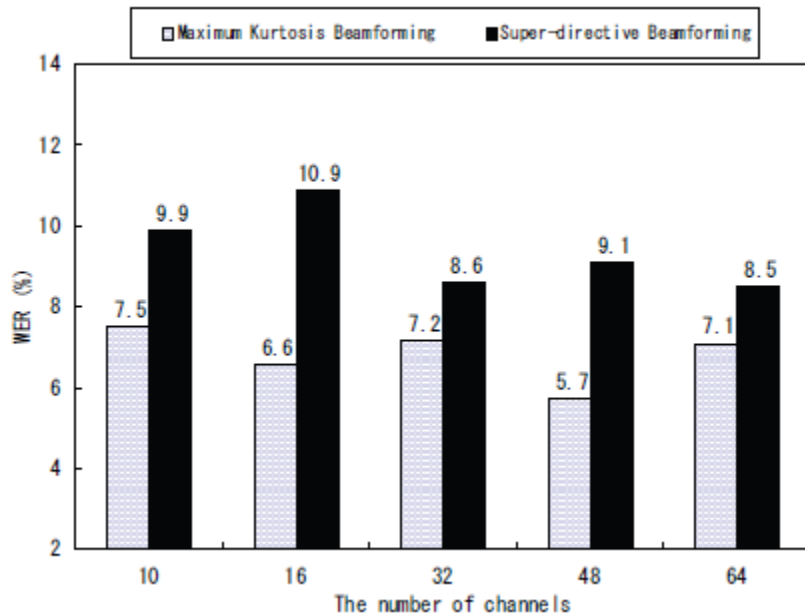
$PAR = (Number\ of\ all\ recognized\ phonemes - Number\ of\ erroneously\ inserted\ phonemes - Number\ of\ erroneously\ deleted\ phonemes - Number\ of\ erroneously\ replaced\ phonemes) / (Number\ of\ all\ phonemes)$

Algorithm	Pass (%WER)		
	1	2	3
Single distant microphone	38.1	19.8	15.4
SD beamforming	24.1	11.3	8.6
MK beamforming	21.8	8.3	5.7
Lapel microphone	19.3	5.9	5.2

Table 1. Word error rates (WERs) for each decoding pass.

Post-filtering (filtr Wienera) – filtr cyfrowy pracujący na widmie sygnału, statystycznie porównuje zarejestrowane dane z czystym sygnałem referencyjnym, następnie dokonuje „odplotu” wyznaczonego szumu







Podsumowanie

- Kryminalistyka (podśluchy)
- Systemy domu inteligentnego
- Rozpoznawanie mowy w wypadku znanej pozycji mówcy
- Zabawki interaktywne (kinect?)

Bibliografia

- „Identyfikacja źródeł aktywności wibroakustycznej maszyn metodą kształtowania wiązki” inż. Jerzy Motylewski, mgr inż. Piotr Pawłowski, mgr inż. Michał Rak, dr inż. Tomasz Zieliński
- „Beamforming Using Uniform Circular Arrays for Distant Speech Recognition in Reverberant Environment and Double-Talk Scenarios” Hannes Pessentheiner
- „Subband Likelihood-Maximizing Beamforming for Speech Recognition in Reverberant Environments” Michael Seltzer, Richard Stern
- „Far-Field Continuous Speech Recognition System based on Speaker Localization and Sub-Band Beamforming” Afsaneh Asefi, Mohammad Javad Taghizadeh, Hossein Sameti
- „Channel selection based on multichannel cross-correlation coefficients for distant speech recognition” Kenichi Kumatani, John McDonough, Jill Lehman, Bhiksha Raj



AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Zastosowanie techniki kształtowania wiązki w celu zwiększenia skuteczności systemów DSR

*Bartłomiej Chojnacki
Maciej Jagoda*

Dziękujemy za uwagę