

# KOMBINACE INFORMACÍ VÍCE-KANÁLOVÝCH NAHRÁVEK

Excel@FIT 2023

FAKULTA  
INFORMAČNÍCH  
TECHNOLOGIÍ

Jan Procházka

Fúze informací více-kanálových nahrávek je testována na úloze verifikace mluvčího. Jako reference jsou zvoleny jedno-kanálové nahrávky a tabulky ukazují relativní zlepšení vůči nim. Jedná se o nahrávky pořízené vzdálenými mikrofony, obsahují velké množství šumu, reverberaci a malý dynamický rozsah. Následující metody se snaží s těmito podmínkami vypořádat.

## ÚROVEŇ SKÓRE

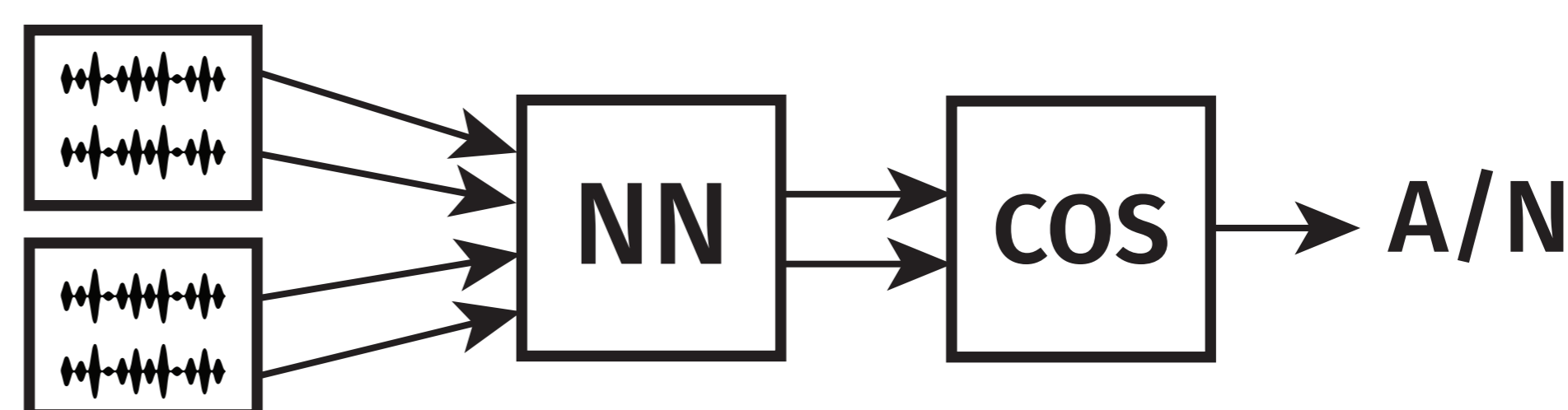
Skóre je reálné číslo, které vypočítá funkce  $\cos$  podobnosti z vektorové reprezentace mluvčího. Pro každý kanál nahrávky je k dispozici právě jedno skóre. Se skóre jsou prováděny matematické operace jako průměr, či hledání maxima.

RETRANSMIT	DEV	EVAL v1
CE	25.84%	70.27%
SRE	9.26%	65.47%
MRE	10.91%	51.64%
MRE HARD	12.39%	37.85%

## ÚROVEŇ EMBEDDINGU

Embeddingy jsou vnitřní reprezentace mluvčího, které vytváří neuronová síť, která je použita pro vyhodnocení verifikace. Embeddingy jsou tvořeny více-rozměrnými vektory, nad kterými jsou prováděny obyčejné vektorové operace jako normalizace, maximum a průměr.

RETRANSMIT	DEV	EVAL v1
CE	7.14%	63.20%
SRE	7.41%	58.59%
MRE	7.11%	54.07%
MRE HARD	13.80%	40.64%



## ÚROVEŇ SIGNÁLU

Při fúzi na úrovni signálu se využívá metody prostorového filtrování (beamforming), výpočty jsou založené na fyzikálních vlastnostech zvukových vln a jejich propagaci ve vzduchu v místnosti. Implementované metody: Delay and Sum (s omezením posunu) a různé varianty MVDR.

RETRANSMIT	DEV	EVAL v1
CE	-3.58%	11.02%
SRE	-16.58%	6.49%
MRE	-16.36%	9.19%
MRE HARD	-8.15%	6.12%

## POROVNÁNÍ METOD

