

Cadenza Workshop

Friday, 8 December 2023

Audio Mixing and Stem Proportions Adjustment

Epri Wahyu Pratiwi, Yu Tsao, Stefano Rini
Academia Sinica, Taiwan
National Yang Ming Chiao Tung, Taiwan

Music Perception in hearing loss person

- Hearing loss can have a major impact on a person's quality of life, making it difficult to communicate, participate in social activities, and enjoy music.
- Hearing aids mostly focused on speech communication, not in music.
 - Problem to pick out the lyrics and melody lines
 - Problem to hear the high frequencies that give the music its richness and detail

Why Music Remixing?

- We want to make listener able to pick up the lyrics
- Methods: Music source separation - isolating instruments and vocals from a mixed audio track.

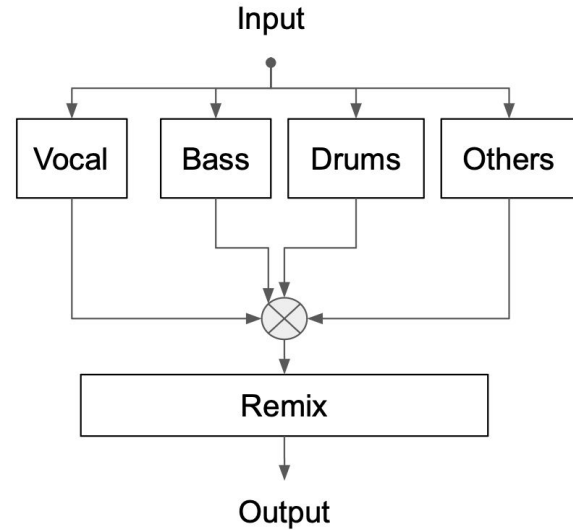
Demux Model

Architecture: A deep learning-based U-Net convolutional structure.

Specialization: Excelling in separating drums, bass, and vocals from a given music track.

Algorithm Overview

1. Source separation
2. Stem presence detection
3. Proportional adjustment
 - Vocals: 60%
 - Bass: 15%
 - Drums: 10%
 - Other: 15%
4. Remixing process



Experimental Results (Task 1)

	score
Baseline	0.2550
Ours	0.1951

	score	left_bass	right_bass	left_drums	right_drums	left_other	right_other	left_vocals	right_vocals	remix_left	remix_right	remix_score
Baseline	0.2550	0.1725	0.1704	0.2231	0.2208	0.3421	0.3440	0.2840	0.2828	0.7031	0.7091	0.7061
Ours	0.1951	0.2000	0.1969	0.1638	0.1449	0.2990	0.1802	0.2091	0.1672	0.2792	0.1555	0.2173

Experimental Results (Task 2)

	score	haaqi_left	haaqi_right
Baseline	0.1256	0.1186	0.1327
Ours	0.1187	0.1122	0.1251

Song Title	Average (%)	
	Baseline	Ours
Al James - Schoolboy Fascination	39.1	46.1*
Angels In Amplifiers - I'm Alright	47.8	43.6
BKS - Too Much	50.1	11.8
Carlos Gonzalez - A Place For Us	48	43.9
Enda Reilly - Cur An Long Ag Seol	55.9	32.9
Forkupines – Semantics	34.8	17.6
Girls Under Glass - We Feel Alright	26.2	23.7
Hollow Ground - Ill Fate	28.1	20.7
James Elder & Mark M Thompson - The English Actor	21.3	28.7*
Little Chicago's Finest - My Own	52.6	34.2
M.E.R.C. Music - Knockout	43	26.3
Moosmusic - Big Dummy Shake	25.4	33.9*
Motor Tapes - Shore	38.1	37.1
Mu - Too Bright	34.1	29.9
Nerve 9 - Pray For The Rain	32.5	38.0*
PR - Happy Daze	54.6	43.1
PR - Oh No	61.4	44.3
Raft Monk – Tiring	13.9	25.5*
Side Effects Project - Sing With Me	42	39.9
Speak Softly - Broken Man	39.9	44.1*
The Doppler Shift – Atrophy	49.9	46.1
The Mountaineering Club - Mallory	62.6	46.3
The Sunshine Garcia Band - For I Am The Moon	39.6	50.9*
Tom McKenzie - Directions	60.5	45.3
Triviul feat. The Fiend – Widow	33.7	38.4*

Conclusion

- Introduction of Demucs model for music source separation with a dynamic proportional remixing step.
- The remixing step adjusts stem proportions based on sound presence, incorporating stem presence detection and proportional adjustment.
- Evaluation using Cadenza Challenge data for Task 1 and Task 2.
- Despite the objective score being below the baseline, the proposed method achieves better subjective scores for multiple songs.
- Emphasis on the positive impact of enhancing vocal information on overall music perception.

Thank you!