

Pengembangan Model SkiM Berbasis *Multi-Head Self-Attention* dan *Transfer Learning* untuk *Speech Separation* Tiga Pembicara Berbahasa Indonesia

Nama Mahasiswa : Hadad Fadilah
NIM : 11221029
Dosen Pembimbing Utama : Bima Prihasto S.Si., M.Si., Ph.D.
Dosen Pembimbing Pendamping : Ramadhan Paninggalih S.Si., M.Si., M.Sc.

ABSTRAK

Teknologi *speech separation* merupakan cabang ilmu *speech processing* yang bertujuan memisahkan suara masing-masing pembicara dari sinyal audio campuran. Sebagian besar penelitian *speech separation* yang ada masih terbatas pada bahasa Inggris, sementara bahasa Indonesia sebagai *Low-Resource Language* masih menghadapi keterbatasan ketersediaan model dan dataset berkualitas tinggi. Selain itu, penelitian yang ada umumnya hanya menangani skenario dua pembicara, padahal kondisi percakapan nyata seperti rapat dan diskusi kelompok umumnya melibatkan tiga pembicara atau lebih. Pelatihan model juga membutuhkan biaya komputasi yang tinggi sehingga menjadi kendala pada lingkungan dengan sumber daya terbatas. Oleh karena itu, penelitian ini bertujuan mengembangkan dan mengevaluasi model *Skipping Memory* (SkiM) yang dimodifikasi dengan mekanisme *Multi-Head Self-Attention* untuk tugas *speech separation* tiga pembicara pada bahasa Indonesia. Penelitian ini membandingkan tiga skenario eksperimen yaitu pelatihan model SkiM sebagai *baseline*, pelatihan SkiM dengan *transfer learning* dari model dua pembicara ke tiga pembicara, dan pelatihan model SkiM yang dimodifikasi dengan *Multi-Head Self-Attention* menggunakan *transfer learning* dari model dua pembicara menuju tiga pembicara. Dataset yang digunakan merupakan dataset sintetis bahasa Indonesia yang dibuat untuk mensimulasikan kondisi percakapan banyak pembicara. Kualitas pemisahan dievaluasi menggunakan metrik SI-SNR. Hasil penelitian menunjukkan bahwa model SkiM mencapai rata-rata SI-SNRi sebesar 18,2157 dB dan STOI sebesar 0,9374, model SkiM dengan *transfer learning* meningkat menjadi 19,4925 dB dan 0,9495, dan model SkiM yang dimodifikasi dengan *Multi-Head Self-Attention* menggunakan *transfer learning* mencapai nilai tertinggi sebesar 19,7753 dB dan 0,9518, atau meningkat sebesar 1,5596 dB pada SI-SNRi dan 0,0144 pada STOI dibandingkan model *baseline*. Hasil ini menunjukkan bahwa kombinasi modifikasi *Multi-Head Self-Attention* dan *transfer learning* efektif meningkatkan kualitas *speech separation* tiga pembicara berbahasa Indonesia.

Kata kunci: *Speech Separation*, SkiM, *Multi-Head Self-Attention*, *Transfer Learning*, bahasa Indonesia