UNIVERSITAS
GADJAH MADA
**Analisis Perbandingan Performa Library Rhubarb dan Azure untuk Real-Time Lipsync pada Sistem Avatar
Interaktif**
Shyra Athaya, Divi Galih Prasetyo Putri, S.Kom., M.Kom, Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

ANALISIS PERBANDINGAN PERFORMA *LIBRARY* RHUBARB DAN AZURE UNTUK *REAL-TIME LIPSYNC* PADA SISTEM AVATAR INTERAKTIF

Shyra Athaya

22/492988/SV/20641

Perkembangan teknologi avatar digital 3D meningkatkan kebutuhan akan sinkronisasi gerakan bibir (*lipsync*) yang realistis untuk mendukung interaksi yang lebih natural. Berbagai *library lipsync* telah dikembangkan untuk tujuan tersebut, di antaranya Rhubarb dan Azure Neural Text-to-Speech. Meskipun keduanya banyak digunakan, perbandingan langsung mengenai efektivitasnya dalam menghasilkan *viseme* untuk animasi avatar 3D masih terbatas. Oleh karena itu, penelitian ini mengkaji efektivitas kedua *library lipsync* tersebut dalam menghasilkan urutan *viseme* yang akurat untuk animasi avatar 3D. Evaluasi dilakukan melalui pengujian kinerja sistem, pengukuran tingkat akurasi *viseme* yang dihasilkan, serta penilaian persepsi pengguna terhadap kualitas animasi yang ditampilkan. Parameter evaluasi yang digunakan meliputi *Real-Time Factor* (RTF) untuk mengukur efisiensi pemrosesan, *Viseme Accuracy Score* (VAS) untuk mengukur kesesuaian *viseme* yang dihasilkan, dan *Mean Opinion Score* (MOS) untuk menilai persepsi pengguna terhadap naturalitas animasi. Hasil penelitian menunjukkan bahwa Rhubarb memiliki efisiensi pemrosesan yang lebih baik dengan rata-rata RTF sebesar 0,2555 dibandingkan Azure Neural Text-to-Speech sebesar 0,4822, meskipun keduanya tetap berada dalam kategori *real-time* (RTF < 1). Dari sisi akurasi, Azure Neural Text-to-Speech unggul dengan rata-rata VAS sebesar 92,81% dibandingkan Rhubarb sebesar 74,03%. Namun, dari sisi persepsi pengguna, Rhubarb justru memperoleh penilaian MOS yang lebih tinggi (3,43) dibandingkan Azure (2,93), menunjukkan bahwa kehalusan transisi antar *viseme* pada Rhubarb dipersepsikan lebih natural meskipun akurasinya lebih rendah. Dengan demikian, penelitian ini diharapkan dapat memberikan referensi bagi pengembang avatar 3D interaktif dalam memilih teknologi *lipsync* yang sesuai dengan kebutuhan sistem. Selain itu, hasil penelitian ini diharapkan dapat memberikan rekomendasi penggunaan Rhubarb dan Azure Neural Text-to-Speech berdasarkan aspek efisiensi, akurasi, dan kualitas visual yang dihasilkan, baik untuk bahasa Inggris maupun bahasa Indonesia.

Kata kunci: animasi *real-time,* avatar 3D*,* a*zure neural text-to-speech, rhubarb library, viseme*

UNIVERSITAS
GADJAH MADA
Analisis Perbandingan Performa Library Rhubarb dan Azure untuk Real-Time Lipsync pada Sistem Avatar Interaktif
Shyra Athaya, Divi Galih Prasetyo Putri, S.Kom., M.Kom, Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

COMPARATIVE PERFORMANCE ANALYSIS OF RHUBARB AND AZURE LIBRARIES FOR REAL-TIME LIPSYNC IN INTERACTIVE AVATAR SYSTEM

Shyra Athaya

22/492988/SV/20641

The development of 3D digital avatar *technology has increased the demand for realistic lip synchronization (lipsync) to support more natural interactions. Various lipsync libraries have been developed for this purpose, including Rhubarb, and Azure Neural Text-to-Speech. Although both are widely used, direct comparisons of their effectiveness in generating visemes for 3D* avatar *animation remain limited. Therefore, this study examines the effectiveness of these two lipsync libraries in producing accurate viseme sequences for 3D* avatar *animation. The evaluation was conducted through system performance testing, measurement of viseme accuracy, and user perception assessment of the resulting animation quality. The evaluation parameters included Real-Time Factor (RTF) to measure processing efficiency, Viseme Accuracy Score (VAS) to measure the accuracy of generated visemes, and Mean Opinion Score (MOS) to assess users' perceptions of animation naturalness. The results show that Rhubarb achieved better processing efficiency, with an average RTF of 0.2555 compared to 0.4822 for Azure Neural Text-to-Speech, although both remained within the real-time category (RTF < 1). In terms of accuracy, Azure Neural Text-to-Speech outperformed Rhubarb, achieving an average VAS of 92.81% compared to 74.03% for Rhubarb. However, in terms of user perception, Rhubarb received a higher MOS rating (3.43) than Azure (2.93), indicating that the smoother transitions between visemes in Rhubarb were perceived as more natural despite its lower accuracy. Therefore, this study is expected to provide a reference for interactive 3D avatar developers in selecting lipsync technology that best suits system requirements. In addition, the results of this study are expected to offer recommendations on the use of Rhubarb and Azure Neural Text-to-Speech based on efficiency, accuracy, and visual quality in both English and Indonesian.*

Keyword: real-time animation, avatar 3D*, azure neural text-to-speech, rhubarb library, viseme*