A COMPARATIVE ANALYSIS OF VARIOUS 3D MESH OPTIMIZATION ALGORITHMS FOR ASSESSING EFFECTIVENESS ON SUSTAINING VIRTUAL VISUAL ILLUSION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

ÜMİT ERONAT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF MODELLING AND SIMULATION

NOVEMBER 2024

A COMPARATIVE ANALYSIS OF VARIOUS 3D MESH OPTIMIZATION ALGORITHMS FOR ASSESSING EFFECTIVENESS ON SUSTAINING VIRTUAL VISUAL ILLUSION

submitted by **ÜMİT ERONAT** in partial fulfillment of the requirements for the degree of **Master of Science in Modelling and Simulation Department, Middle East Technical University** by,

Date: 29.11.2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: ÜMİT ERONAT

Signature :

ABSTRACT

A COMPARATIVE ANALYSIS OF VARIOUS 3D MESH OPTIMIZATION ALGORITHMS FOR ASSESSING EFFECTIVENESS ON SUSTAINING VIRTUAL VISUAL ILLUSION

ERONAT, ÜMİT

M.S., Department of Modelling and Simulation

Supervisor: Prof. Dr. Saniye Tuğba TOKEL

November 2024, 106 pages

3D modeling is essential in fields like architecture, engineering, and virtual reality, but high-resolution 3D models often demand significant data and computational resources. Mesh simplification algorithms address this issue by reducing model complexity while maintaining visual and structural fidelity. In this study, it is aimed to make a comparative analysis of three mesh simplification algorithms to evaluate how effectively these algorithms perform while still preserving high visual fidelity, specifically by maintaining a profound auditory-visual illusion, referred to as McGurk effect. Since the McGurk effect is a subjective phenomenon, a voluntary user study was conducted to gather data to make the comparison. In the scope of this study, the user study was conducted with 42 participants, of whom 29 passed the preliminary test and contributed usable data for the analysis. Participants were presented with a series of video questions featuring different simplified variations of a head model used for lip-sync animation, and asked to identify what they perceived. This head model was simplified at different levels of vertex reduction using the compared mesh simplification algorithms. The findings showed that different mesh simplification algorithms can exhibit significant variations in their ability to sustain the McGurk effect, depending on the syllable. The findings also provide valuable insights for model designers working with the McGurk effect and offer guidance for future research utilizing mesh simplification algorithms in 3D models for McGurk effect studies.

Keywords: 3D mesh simplification, visual illusion, McGurk effect, comparative analysis

ÖZ

ÇEŞİTLİ 3B ÖRGÜ OPTİMİZASYON ALGORİTMALARININ YANILSAMA ALGILAMASINA ETKİLERİNİN ANALİZ EDİLEREK KIYASLANMASI

ERONAT, ÜMİT
Yüksek Lisans, Modelleme ve Simülasyon Bölümü
Tez Yöneticisi: Prof. Dr. Saniye Tuğba TOKEL

Kasım 2024, 106 sayfa

3B modelleme, mimarlık, mühendislik ve sanal gerçeklik gibi alanlarda önemli bir yere sahiptir, ancak yüksek çözünürlüklü 3B modeller genellikle önemli miktarda veri ve hesaplama kaynağı gerektirir. Bu sorunun çözümünde, ağ örgü basitleştirme algoritmaları, görsel ve yapısal sadakati korurken model karmaşıklığını azaltarak etkili bir çözüm sunar. Bu çalışmada, üç farklı ağ sadeleştirme algoritmasının karşılaştırmalı bir analizi yapılması ve bu algoritmaların, McGurk etkisi olarak adlandırılan derin bir işitsel-görsel yanılsamayı koruyarak görsel bütünlüğü ne kadar etkin bir şekilde muhafaza ettiğinin değerlendirilmesi amaçlanmıştır. McGurk etkisi öznel bir olgu olduğundan, karşılaştırmayı gerçekleştirmek adına veri toplamak üzere gönüllü bir kullanıcı çalışması yürütülmüştür. Bu çalışma kapsamında, 42 kişi katılım sağlamış, bunlardan 29'u ön testi geçmiş ve analiz için kullanılabilir veri sağlamıştır. Katılımcılara, dudak senkronizasyonlu animasyonlar için kullanılan bir kafa modelinin farklı basitleştirilmiş varyasyonlarını içeren bir dizi video sorusu sunulmuş ve ne algıladıklarını tanımlamaları istenmiştir. Bu kafa modeli, karşılaştırılan ağ örgü sadeleştirme algoritmaları kullanılarak farklı yüzey azaltma seviyelerinde basitleştirilmiştir. Bulgular, farklı ağ sadeleştirme algoritmalarının McGurk etkisini sürdürme yeteneklerinde, heceye bağlı olarak önemli farklılıklar gösterebileceğini ortaya koymuştur. Çalışmanın sonuçları, McGurk etkisiyle çalışan model tasarımcıları için değerli bilgiler sunmakta ve bu tür algoritmaların McGurk etkisi araştırmalarında 3B modellerde nasıl kullanılabileceğine dair gelecekteki çalışmalara rehberlik etmektedir.

Anahtar Kelimeler: 3B model sadeleştirme, görsel ilüzyon, McGurk etkisi, karşılaştırmalı analiz

To my beloved parents and my little flower, to my altruistic companion 35KUM01, who carried me all those years, and to life, for finally letting me see the end of this...

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Prof. S. Tuğba Tokel, for her guidance, feedback and insight throughout my research. Her knowledge and experience were instrumental in the completion of this thesis.

I am also thankful to all of the distinguished professors in the examining committee; Prof. Alptekin Temizel and Prof. Hasan Çakır, for their invaluable feedback, understanding and appreciation.

I also appreciate all the participants and their thoughtfulness, dedication and contributions, which made this study possible.

Finally, the completion of this thesis could not have been possible without my beloved family, and of course, Zeynep Çakır, for without her endless love, support and understanding I would surely be lost. She is one of the biggest reasons I was able to see the end of my work, and I will be forever grateful to her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

ALGORITHMS

LIST OF ABBREVIATIONS

3D	Three-dimensional
CAD	Computer-Aided Design
QEM	Quadric Error Metrics
CGI	Computer-Generated Imagery
XML	Extensible Markup Language
LOD	Level of Detail
ANOVA	Analysis of Variance

CHAPTER 1

INTRODUCTION

This chapter gives a brief background of this study, before continuing with a paragraph about the main problem that drives the study. Later, the chapter mentions about the significance of this study and how its contributions. Finally the chapter will end after stating the research questions and giving commentaries about implications and possible future work regarding this research.

1.1 Background of the Study

Computer games, animated films and virtual reality are some of the areas where virtual 3D (three-dimensional) model design is effectively utilized. In such areas the visual realism of a scene, or how exact it reflects the idea in the designer's mind, is of critical importance. Knowing this, 3D model designers aim to achieve a high visual-fidelity content using various modeling tools; such as Blender, Autodesk 3ds Max, Maya, etc. which are some of the industry's prevalent tools. When creating a visually pleasing model, a designer must also take into account one crucial aspect of the production; which is the storage cost of each produced asset. The effect of this cost may, for example, impact the storage space that is consumed in the disks of the computers on which the 3D models are produced, or increase the bandwidth of a network infrastructure which these assets are transmitted through. Having a firm grasp of this issue, every designer aspires to increase the visual fidelity of the produced models with as little expenditure as possible, consequently achieving an increase in cost-effectiveness.

To elaborate more on this cost-effectiveness, it is important to understand what a virtual 3D model is actually composed of and how it affects storage consumption. Essentially, every 3D model consists of points in 3D space. These points are most commonly known as vertices, and they are represented in Cartesian coordinates. Considering that every coordinate must occupy a certain space on a storage medium, so that the model can be reused using this data, as the number of vertices increases in a model, the data size of the model also increases proportionally. This "vertex count-model size" correlation is referred to as the "digital footprint" in the following sections.

1.2 Statement of the Problem

As the industry is seeing storage limitations [6], designers that work in areas such as computer games, animated films, or academic studies need to consider the increasing costs of designing extremely detailed models. This often constitutes a substantial problem that becomes an obstacle for achieving the desired output. Because, as the digital footprint of a model gets larger, it becomes increasingly infeasible to render the 3D content [7]. But the designers cannot constantly refrain from adding more details so that the storage consumption is kept minimal; as that would mean high visual fidelity becomes less and less feasible. Therefore, producing optimal 3D assets becomes a non-trivial challenge. One of the main approaches to overcome this challenge is producing assets with a cost-efficient (optimized) digital footprint. To elaborate further, it is imperative to delve deeper about presenting cost-efficiency methods.

A cost-efficiency method is meant to remarkably reduce the storage consumption of a 3D model so that it takes up substantially less space while at the same time the visual detail in the model is kept at a satisfactory level. The means to reduce this consumption usually involves reducing the main source of data, which is the number of vertices in the model. However, achieving such a goal without sacrificing high visual fidelity is not a trivial problem. The vertices cannot be randomly reduced, or otherwise the overall look and feel of the model can be lost. Thus, certain techniques, which are often complex, are utilized in order to reduce the number of vertices without sacrificing visual fidelity. Reducing the vertex count this way is often referred to as model optimization or mesh simplification. Various mesh simplification algorithms have been developed. However, the cost effectiveness of each algorithm can be comparatively different [8], and making a definitive better-worse comparison between different algorithms is not in the scope of this study. The main aim of this study is to put forward a novel way to comparatively analyze different mesh simplification algorithms using a very specific criteria. The problem, then becomes, whether a specific audio-visual illusion, the McGurk effect, can still be observed at certain vertex reduction levels of different algorithms. This approach will then provide a guidance to those who study the same illusion in virtual 3D environments, and to those who plan on using mesh simplification algorithms for such projects.

1.3 Research Questions

The main idea behind this study is to observe whether it is possible to achieve a comparative analysis between three specific mesh simplification algorithms using the specific methodology of this study. This analysis is based on how cost-efficient the compared algorithms can be, relative to each other, by observing the correlation between reduction in storage cost vs. maintaining the McGurk illusion. As such, the analysis is intended to be kept strictly objective, relying on the gathered data from a participated user study. Thus, the following question form as the main driver of this study:

- Is there a significant difference between three different mesh simplification algorithms with two vertex reduction levels (35% and 70%) on the maintenance of the McGurk effect in 3D head models used for lip-sync animation?

1.4 Significance of the Study

This study combines the aspects of 3D modelling, generating different resolutions of meshes via various mesh simplification algorithms and the changes of audio-visual perception under the McGurk effect, with an aim to provide guidance to experts who carry out modeling and design activities on the model simplification algorithms they use in their fields. In this way, making an efficiency comparison between the three most frequently used simplification algorithms in modeling fields will be a time and effort saver for these people. The use of a visual illusion as a medium to gather statistical data and to comparatively analyze any collection of mesh simplification algorithms at any time emerges as the novelty of the study.

1.5 Definition of Important Terms

3D: 3D, or three-dimensional, refers to an object or space that has three dimensions: width, height, and depth. Unlike two-dimensional (2D) objects, which only have width and height (like a flat image or drawing), 3D objects can be viewed and interacted with from multiple angles, giving them a more lifelike and realistic appearance.

Vertex: A vertex is a fundamental element in the representation of a 3D object within computer graphics and computational geometry. It is essentially a point in 3D space, defined by its coordinates. Vertices are used to construct more complex geometric shapes such as edges, faces, and ultimately, entire 3D models.

Mesh: A mesh is a collection of vertices, edges, and faces that defines the shape of a 3D object in computer graphics and computational geometry. Meshes are fundamental in the creation and representation of 3D models and are used extensively in various applications, including video games, simulations, animation, and scientific visualization.

Level-of-Detail (LOD): It is a technique used in computer graphics, particularly in 3D modeling and rendering, to manage and optimize the complexity of a scene. The main idea behind LoD is to reduce the number of polygons in a 3D model as the model moves further away from the viewer or as it becomes less significant to the overall scene. This process helps to maintain high performance and efficiency in rendering without noticeably compromising visual quality. The concept was introduced by James Clark, in his seminal 1976 paper titled "Hierarchical Geometric Models for Visible Surface Algorithms."

1.6 Organization of the Thesis

Given the introduction to the study, the thesis proceeds immediately with the related work section; outlining various 3D mesh optimization algorithms, the 3D modeling software used in this study, as well as briefly explaining the McGurk Effect and the contemporary academic work on these subjects. Later, the methodology section lays out the steps for the proposed method of this study, which then continues with the detailed explanation of the user study that was conducted to realize solid work for this study. Lastly, the thesis concludes with two consecutive sections, the former of which presents the results of the conducted user study, while the latter section concludes the thesis with final thoughts, considerations and future work.

CHAPTER 2

LITERATURE REVIEW

To better understand the cost-effectiveness relation between 3D models and mesh simplification algorithms, this section gives preliminary information by reviewing existing literature on 3D modeling, mesh optimization and highlighting their applications. And since a specific visual illusion contributes to the evaluation and analysis phase of this study, this chapter also refers to that illusion, which is the McGurk effect, its impact on perception, and how it can be used as a comparative framework.

2.1 3D Modeling and Essential Concepts

3D modeling is a multifaceted field within computer graphics that involves the creation of three-dimensional representations of objects, environments, and characters. This process is fundamental to various industries, including animation, gaming, virtual reality, and engineering [9]. Below are some of the key concepts of 3D modelling:

Geometry refers to the mathematical study of shapes and their properties. In 3D modeling, geometry defines the structure of a model through a mesh and its main components; vertices, edges, and faces, forming the basic building blocks known as polygons. Common geometric primitives, such as cubes, spheres, and cylinders, serve as the starting point for more complex models.

Topology involves the arrangement and flow of vertices, edges, and faces on a 3D surface. Proper topology ensures that a model deforms correctly during animation and maintains its structural integrity.

Rendering is the process of converting a 3D model into a 2D image or animation. Rendering algorithms calculate the interactions of light with the model's surfaces to produce the final visual output.

In addition to these three important concepts, there is also *texturing*, which is an application of images or patterns to a 3D model's surface to simulate various materials, and *lighting*, which is an application of real-world illumination to create depth, shadows and highlights.

2.1.1 Model vs Mesh

In the context of 3D graphics and geometric modeling, the terms "mesh" and "model" are related but distinct concepts. A *mesh* is a collection of vertices, edges, and faces that defines the shape of a 3D object. It is a geometric representation used to approximate the surface of an object. A 3D *model* is a broader term that encompasses not just the geometric shape (mesh) but also additional attributes and data that define a 3D object in a digital space. Meshes are primarily used to represent the external surface of 3D objects, comprising of vertices, edges and faces; whereas a model provides a complete representation of a 3D object, including its visual appearance, physical behavior, and how it interacts with other objects. A model may have textures, materials, rigging and animations, even physics related data, all of which resides in a hierarchical organization. Through this hierarchical organization, 3D models can accurately depict detailed surfaces and volumes, facilitating tasks such as rendering, simulation, and analysis in various fields including computer graphics, virtual reality, and engineering. Regarding cost-effectiveness, components of a mesh (vertices, edges and faces) are substantially more important than a model's other aspects, since those components contribute the most when it comes to consuming disk space.

2.1.2 Components of a 3D Mesh

Below are the definitions of vertices, edges and faces. An introductory image of these definitions is provided in Figure 1.

1. Vertices (Points): A vertex in the context of 3D modeling and computer graphics refers to a fundamental point in three-dimensional space. It is defined by its coordinates, typically represented as (x, y, z) in a Cartesian coordinate system. Vertices serve as the basic building blocks of 3D models, marking the precise locations where edges meet and determining the shape and structure of geometric primitives such as polygons and curves. They play a crucial role in defining the spatial properties and connectivity of a 3D model, facilitating accurate representation and manipulation in digital environments.

2. Edges (Lines): Edges connect pairs of vertices, forming straight or curved lines that delineate the boundaries between adjacent faces. They represent the structural framework of the mesh and define the geometric relationships between vertices.

3. Faces (Polygons): Faces are planar surfaces defined by three or more vertices connected by edges. Common types of polygons used in 3D meshes include triangles (three vertices) and quadrilaterals (four vertices). Faces enclose the volume and surface area of the object, representing its visible or textured surfaces.

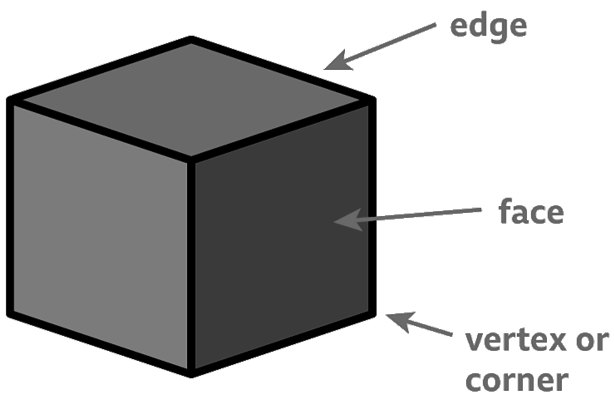


Figure 1: Introductory image of what a vertex, an edge and a face is. (Available at: `https://ichef.bbci.co.uk/images/ic/976xn/p08j38s9.png`)

2.1.3 Designing and Modeling

A designer creates a 3D model through a structured process involving several stages of digital manipulation and visualization. Initially, the designer conceptualizes the model's form and function based on project requirements and artistic vision. This conceptual phase may involve sketches, reference materials, or digital concept art to outline the model's design parameters and aesthetic qualities.

Subsequently, the designer transitions to the modeling phase, where digital tools such as computer-aided design (CAD) software or specialized 3D modeling applications are utilized. In this phase, the designer employs geometric primitives, such as vertices, edges, and faces, within a Cartesian coordinate system to construct the model's basic structure. Techniques such as extrusion, scaling, and sculpting refine the geometry, shaping intricate details and surface textures to accurately represent the intended object or scene.

Throughout the modeling process, the designer iteratively refines the 3D model, adjusting proportions, refining surfaces, and optimizing geometry to achieve desired levels of realism or functionality. This iterative refinement may involve techniques such as polygonal modeling, spline modeling, or digital sculpting, depending on the complexity and artistic requirements of the model.

Once the modeling phase is complete, the designer proceeds to the texturing and material assignment stage. Here, surface materials, colors, textures, and shaders are applied to enhance visual realism and simulate physical properties of the model. Techniques such as UV mapping, procedural texturing, and image-based texturing are employed to map 2D textures onto the 3D surfaces effectively.

Following texturing, the model enters the lighting and rendering phase, where virtual lighting setups and rendering settings are configured to illuminate and visualize the model realistically. Advanced rendering techniques such as ray tracing or global illumination may be utilized to simulate light interactions and shadow effects, further enhancing visual fidelity.

Finally, the designer evaluates and iterates on the completed 3D model, ensuring that it meets project specifications, aesthetic goals, and technical requirements. Throughout this process, collaboration with stakeholders, feedback incorporation, and adherence to industry standards contribute to producing a finalized 3D model that effectively communicates the designer's vision and serves its intended purpose in various applications such as animation, visualization, simulation, or manufacturing. As a last step, the designer stores the produced models in a storage disk using various kinds of 3D model data representations, most notably known as file formats.

2.1.4 3D Model Data Representation and File Formats

3D model data are represented in several ways depending on the context and the specific requirements of the application. Although these representations offer different aspects to embedding the data of a model, and there have been studies to research different techniques on this matter [10], they are all collectively needed together to form a model. Here are some common representations:

- Geometric Representation: This involves defining the shape and structure of the 3D model using geometric primitives such as vertices, edges, and faces. This representation is fundamental for rendering and visualizing the 3D model.

- Surface Representation: Describes the outer boundary or surface of the 3D object. Surfaces can be represented using mathematical equations (parametric surfaces), or as a collection of discrete points (point clouds) or polygons (meshes). This representation is crucial for applications requiring accurate surface rendering and manipulation.

- Material and Texture Representation: Includes information about the visual appearance of the 3D model, such as color, texture, transparency, and reflectivity. Materials and textures can be stored as image files (texture maps), procedural algorithms, or mathematical descriptions, allowing for realistic rendering and simulation of surface properties.

- Structural Representation: Specifies the hierarchical structure and organization of components within the 3D model. This includes grouping objects into hierarchies (parent-child relationships), defining transformations (translation, rotation, scaling), and organizing complex assemblies or scenes. Structural representation facilitates animation, simulation, and interaction in virtual environments.

- Metadata and Annotations: Additional data associated with the 3D model, such as authorship information, creation date, licensing terms, or annotations describing specific features or characteristics. Metadata enhances model management, collaboration, and interoperability across different software and platforms.

- File Formats: 3D models are often stored and exchanged using specific file formats (e.g., OBJ, STL, FBX) that encapsulate geometric, surface, material, and structural data. These formats vary in complexity, capabilities, and compatibility with different software applications and industry standards. Various file formats are used to store and exchange 3D models, each tailored to specific applications, features, and requirements. Common types of 3D model file formats include:

– OBJ (Wavefront OBJ): A versatile format supporting geometry, materials, textures, and basic animations. Widely used in 3D graphics and modeling software.

– STL (Stereolithography): Primarily used for 3D printing, storing the surface geometry of a 3D object as a collection of triangles.

– FBX (Filmbox): Developed by Autodesk, supports animation, materials, textures, and scene hierarchy. Used in game development, animation, and virtual reality.

– Collada (DAE): An XML-based format for exchanging digital assets between various 3D software applications, supporting geometry, materials, textures, and animations.

– PLY (Polygon File Format): Stores a description of a single 3D surface mesh that consists of a set of vertices and triangular faces. Often used in scientific and medical applications.

The vertices are stored in each file format in a unique way. For example, in a Collada file (see Figure 2), the vertex data is stored as a string of numbers in an XML file format. Each format has its strengths and limitations in terms of file size, compatibility, and functionality, catering to different aspects of 3D modeling, visualization, and production workflows [11]. However, whichever format is used, the size of the data is always directly proportional to the number of vertices in the model's mesh. Therefore, it is crucial to reduce this number in order to achieve a remarkable level of effectiveness. For that purpose, mesh simplification is indeed a feasible way as the main purpose of such an algorithm is to reduce the number of vertices in a mesh.

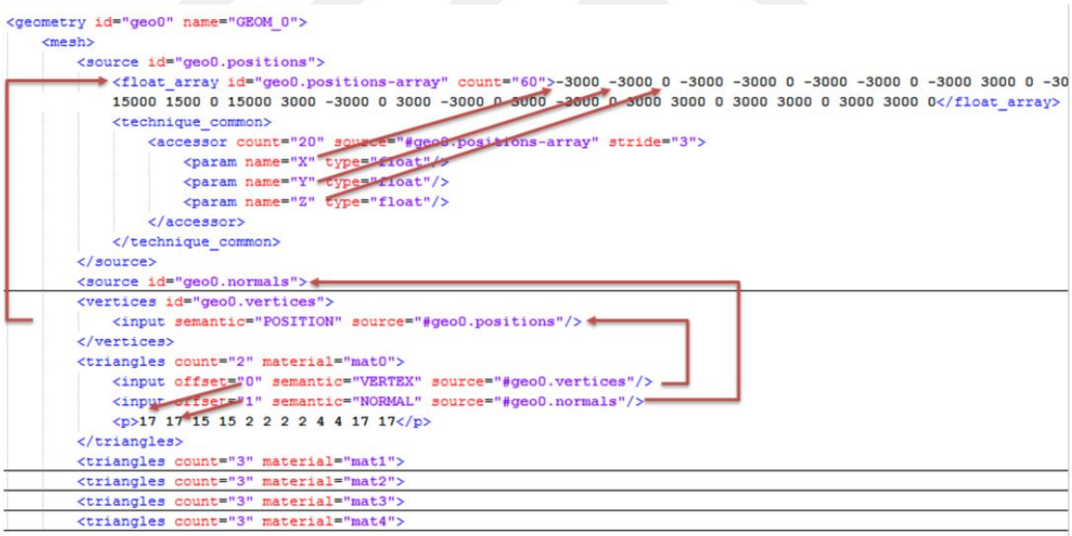


Figure 2: An example Collada (.dae) file. Notice the semantics of the file format and how each vertex is represented as a sequence of real values. (Available at: `https://i.sstatic.net/jDdJE.jpg`)

2.2 Mesh Simplification

3D mesh simplification is a computational technique used in computer graphics and geometric modeling to reduce the complexity of a three-dimensional mesh while preserving its essential shape and appearance characteristics. The goal is to create a simplified version of the mesh that requires fewer computational resources and storage space without significantly compromising visual fidelity or geometric details. This section reviews seminal approaches and methodologies aimed at improving the efficiency and fidelity of triangular meshes.

This technique involves algorithms that selectively remove vertices, edges, or faces from the original mesh based on criteria such as geometric error, curvature, or visual importance. By intelligently reducing the number of elements in the mesh, applications can achieve faster rendering times, efficient transmission over networks, and improved performance in interactive environments like virtual reality or real-time simulations [12]. An example of such reduction is shown in Figure 3 and Figure 4.

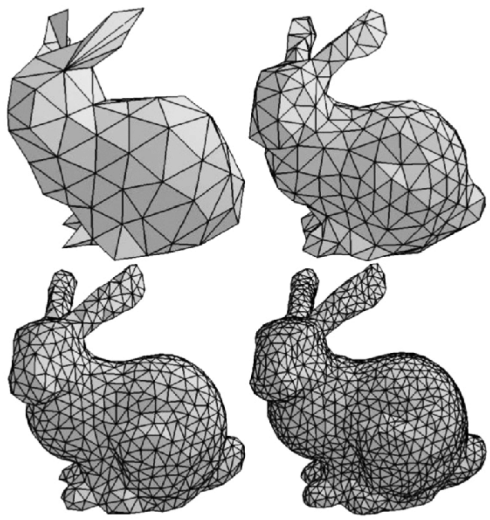


Figure 3: 3D mesh-triangles with different resolutions (3D Modelling for programmers. [1])

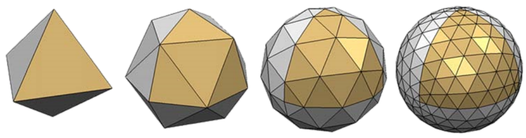


Figure 4: An example of surface subdivision. It can be noticed that at each step the number of faces are increased by 4, while the edges are halved in distance. *Image sourced from the proceedings of ACM SIGGRAPH 2006 [Botsch et al., 2006]* [2]

2.2.1 Various Algorithms

The field of polygonal mesh simplification has a history spanning over two decades, originating in 1976 when James Clark first introduced the concept of levels of detail (LOD) [13]. Mesh simplification aims to reduce the complexity of 3D models while preserving essential geometric and topological properties. There are many algorithms that can be used, and many more in the current conjecture are researched progressively. Regardless of the chosen technique, as the impact of the algorithm is increased, the LOD decreases.

The following subsections briefly underlines some key notes for a few algorithms that are commonly used in the industry. In addition to these algorithms, Cignoni et al. also conducted a comparative study evaluating various mesh simplification algorithms [8]. Their analysis highlighted edge-collapse operations and their trade-offs between computational efficiency and preservation of surface details.

2.2.1.1 Quadric Error Metrics

Developed by Michael Garland and Paul Heckbert in 1997 [14], the Quadric Error Metrics (QEM) algorithm is a popular and foundational mesh simplification technique used to reduce the number of polygons in a 3D model while preserving its appearance as much as possible. This method is noted for achieving high-quality simplifications with minimal distortion.

The QEM algorithm simplifies a mesh by iteratively collapsing edges onto the plane of the mesh. Firstly, for each vertex, a 4x4 quadric matrix is computed, representing the error associated with collapsing or moving that vertex. Secondly, for each edge in the mesh, the cost of collapsing that edge is calculated to minimize the error introduced to the mesh. Next, this error is quantified using a "quadric" which measures the distortion caused by collapsing an edge. It is derived from the sum of squared distances between the vertices of the original mesh and the plane that best fits these vertices. Then, the collapse on the edge that introduces the least error is performed, meaning merging the two vertices at the ends of the edge into a single vertex, and updating the mesh accordingly. Lastly, the error quadrics of the vertices affected by the collapse is updated. This update involves adjusting the quadric matrices to reflect the new vertex positions and the updated mesh topology. This whole process is repeated until the desired level of simplification is achieved or until a specified number of edges have been collapsed.

The QEM algorithm is effective not only at preserving the visual fidelity of the meshes with detailed features, but also handling large meshes. Note that for very large meshes, the computation of edge costs and the updating of quadrics can be computationally intensive.

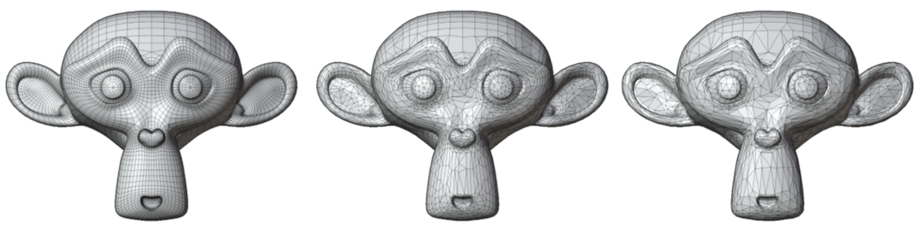


Figure 5: An example illustration of a comparison between different variants of the same mesh using an implementation of the QEM algorithm.

2.2.1.2 Triangular Mesh Simplification Based on Surface Angle

Triangular Mesh Simplification based on Surface Angle is another approach used to simplify 3D meshes and it is proposed by Li et al. [15].

This technique focuses on the surface angles between adjacent faces at a shared edge in a mesh to determine which triangles can be removed or merged with minimal impact on the overall shape, meaning focusing in areas where the angle between adjacent faces is small. High angles typically denote sharp features, while low angles indicate smoother transitions.

This algorithm initiates by calculating the angle between the two adjacent faces that share the edge. This is performed by using the dot product of the normal vectors of the faces, for each edge in the mesh. Then, the simplification criteria is determined by setting a threshold angle: If the angle between adjacent faces is below this threshold, the triangles are considered less important and are candidates for removal or merging. Next, the simplification process is performed for each candidate edge by calculating the cost of collapsing the candidate edge and updating the mesh by removing the collapsed edge, merging the vertices, and adjusting the connectivity of the affected faces. This whole process is repeated until the desired level of simplification is achieved or until further simplification would significantly impact the mesh quality.

This technique effectively preserves sharp features and edges by focusing on surface angles, which is useful for retaining important geometric details.

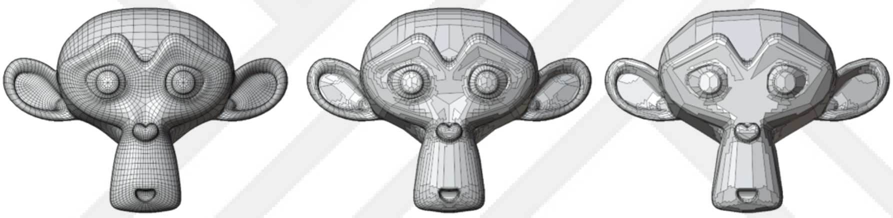


Figure 6: An example illustration of a comparison between different variants of the same mesh using an implementation of the Triangular Mesh Simplification Based on Surface Angle algorithm.

2.2.1.3 Hierarchical Face Clustering on Polygonal Meshes

Hierarchical Face Clustering on Polygonal Meshes is a powerful technique presented by Garland et al. [16] for managing the complexity of 3D meshes while preserving important details, making it suitable for applications where both accuracy and performance are critical. It organizes and simplifies a 3D mesh by clustering faces hierarchically, meaning that it is first applied to large clusters and then refined in smaller sub-clusters.

This algorithm starts by grouping faces that are similar in terms of their properties (texture, normal direction, etc.) and creating initial clusters of faces based on similarity. Secondly, a hierarchy of clusters is constructed by a tree-like structure where the top level clusters are large and encompass many faces, while lower levels represent smaller, more detailed clusters. Then, from bottom to top, these clusters are merged to form larger clusters in order to progressively simplifying the mesh while maintaining important structural details. Next, each large cluster is simplified by reducing the number of faces by edge collapse or vertex clustering techniques within the cluster. It is important to ensure that simplification within a cluster does not significantly degrade the visual quality or structural integrity of the mesh. After simplifying large clusters, the simplification in smaller sub-clusters is refined, and the structure is adjusted to these simplifications or refinements. At the end, the final simplified mesh is constructed by combining the simplified clusters. Note that the final mesh must show consistency in terms of topology and appearance.

This hierarchical approach allows for controlled and gradual simplification that maintains the model's important features while reducing complexity. Also, this approach can be adapted to various levels of detail, making it suitable for different applications and requirements.

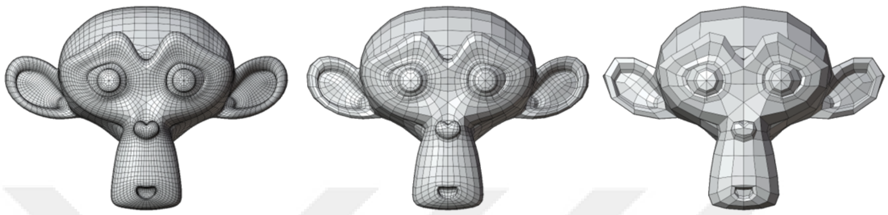


Figure 7: An example illustration of a comparison between different variants of the same mesh using an implementation of the Hierarchical Face Clustering on Polygonal Meshes algorithm.

In summary, the evolution of 3D mesh optimization techniques highlights a multidisciplinary effort to achieve computational efficiency and visual fidelity across diverse applications. Continued exploration of novel methodologies integrating computational geometry, machine learning, and human-computer interaction promises to address emerging challenges and unlock new opportunities in the field.

2.3 Modeling Software

In contemporary design and simulation practices, the utilization of advanced 3D modeling software has become pivotal. These software tools facilitate the creation, manipulation, and visualization of intricate three-dimensional models, enabling designers, engineers, and artists to manifest creative concepts into tangible digital representations. Such software, exemplified by industry-standard platforms like Autodesk Maya, Blender, and SolidWorks, offers a diverse array of tools ranging from parametric modeling and sculpting to animation and virtual reality integration.

The significance of 3D modeling software extends beyond mere representation, encompassing crucial roles such as in prototyping and engineering analysis [17], as well as immersive virtual environments [18]. By providing robust capabilities for precision modeling, material simulation, and real-time rendering, these tools empower professionals across diverse disciplines to innovate and refine designs with unprecedented accuracy and efficiency. Moreover, their integration within contemporary workflows underscores their indispensable role in shaping modern design methodologies and enhancing collaborative endeavors.

2.3.1 Blender

One example to 3D modeling software is Blender, which is developed and maintained by the Blender Foundation, a non-profit organization based in the Netherlands. Being an open-source 3D modeling and animation software, it has emerged as a pivotal tool in contemporary digital content creation and visualization. Developed and maintained by a global community of developers, Blender exemplifies the convergence of accessibility, versatility, and technological innovation in the realm of computer-generated imagery (CGI). Its comprehensive suite of features spans from polygonal modeling and sculpting to rigging, animation, and rendering, thereby accommodating diverse creative workflows across animation studios, game developers, and visual effects professionals. A screenshot from the software is provided in Figure 8.

Blender is a comprehensive and versatile 3D modeling software widely utilized in various disciplines such as computer graphics, animation, and virtual reality. Having differences with other modeling software, such as 3D Studio Max [19], it is known for its robust feature set, which includes advanced modeling tools, sculpting capabilities, rendering engines, and animation functionalities. Blender's open-source nature fosters a collaborative environment for developers and artists to innovate and customize workflows, making it a prominent tool in both educational and professional settings.

Figure 8: A screenshot from Blender. (Available at: `https://www.blender.org/`)

Blender provides several tools and techniques that aid in mesh simplification, each serving specific purposes to reduce polygon count and optimize 3D models while maintaining their essential visual and structural integrity. Here's a various number of algorithms that are relevant and studied in this thesis are implemented by Blender:

1. **Decimate Geometry:** The Decimate Modifier in Blender provides a straightforward approach to reducing polygon count. It offers different algorithms (such as Collapse and Planar) to simplify the mesh while preserving important features. Users can adjust parameters like the ratio of reduction or angle limits to control the degree of simplification. This tool is effective for quickly reducing the complexity of a mesh without significantly altering its overall shape.

2. **Limited Dissolve:** Limited Dissolve is another tool in Blender used for simplifying meshes by merging vertices and edges based on specified criteria. It selectively removes unnecessary

geometry, such as coplanar faces or edges with low angles, to create a cleaner and more efficient mesh structure. Limited Dissolve helps in reducing polygon count while maintaining surface continuity and reducing the complexity of intricate details.

3. **Un-Subdivide:** Un-Subdivide is a tool designed to reverse the subdivision process applied to a mesh. It allows users to iteratively reduce the level of subdivision applied to a model, effectively simplifying the geometry by merging vertices and collapsing edges back to their original state. This tool is useful for refining high-polygon models created through subdivision surface modeling, providing control over the level of detail retained in the simplified mesh.

2.4 Visual Illusions and Altering Perception

A visual illusion, also known as an optical illusion, is a phenomenon in which the perception of a visual stimulus differs from objective reality. This occurs when the brain processes visual information in ways that create discrepancies between the actual physical properties of the stimulus and the perceived image. Visual illusions can reveal a lot about how the human visual system works, including the mechanisms of perception, cognition, and the interpretation of sensory information [20]. An example of a visual illusion is the McGurk Effect, which can be considered a type of visual illusion, though it is more accurately described as an audiovisual illusion.

2.4.1 McGurk Effect

The McGurk effect, first described by Harry McGurk and John MacDonald in 1976 [21], is a perceptual phenomenon that illustrates the interaction between auditory and visual stimuli in speech perception. This effect demonstrates how conflicting visual and auditory components can lead to a third, distinct perception that differs from either the visual or auditory input alone. The original study provided a critical insight into the multifaceted nature of human perception, emphasizing the integrative processes that occur in the brain when interpreting sensory information.

The discovery of the McGurk effect arose from an experiment designed to explore the developmental aspects of speech perception in children. McGurk and MacDonald presented participants with video recordings of a person articulating one phoneme while a different phoneme was dubbed over the audio. For example, the visual component might display the lip movements for "ga," while the audio component would present the sound "ba." Participants often reported hearing a third sound, such as "da," which was neither the visual nor the auditory input but rather a fusion of the two.

This groundbreaking finding was published in the journal Nature, and it has since become a cornerstone in the fields of cognitive psychology, linguistics, and neuroscience. The McGurk effect has profound implications for understanding the complexity of sensory integration and the mechanisms underlying speech perception. Although the illusion was first discovered in 1976, there are still contemporary research done, such as the investigation of how visual components impact this effect, to research how to improving the effect and its realism [22]. Or, such as this research about approaching the illusion with a computational analysis on newer and more modern frontiers, which is mainly about "examining the influence of the temporal presence of sensory cues on the audio-visual percept and exploring how computational models can effectively incorporate the inherent uncertainty associated with audio-visual percepts to predict confidence." [23].

While traditional visual illusions involve discrepancies between visual perception and physical reality, the McGurk effect extends this concept to the integration of visual and auditory information. The illusion lies in the altered auditory perception caused by the incongruent visual stimuli:

- Multisensory Integration: The McGurk effect exemplifies how the brain integrates visual and auditory information to produce a unified perception. The illusion occurs because the visual input modifies the interpretation of the auditory signal.
- Perceptual Overlap: Similar to visual illusions where visual context affects perception (e.g., the Müller-Lyer illusion), the McGurk effect shows how visual context (lip movements) influences auditory perception (heard phonemes).

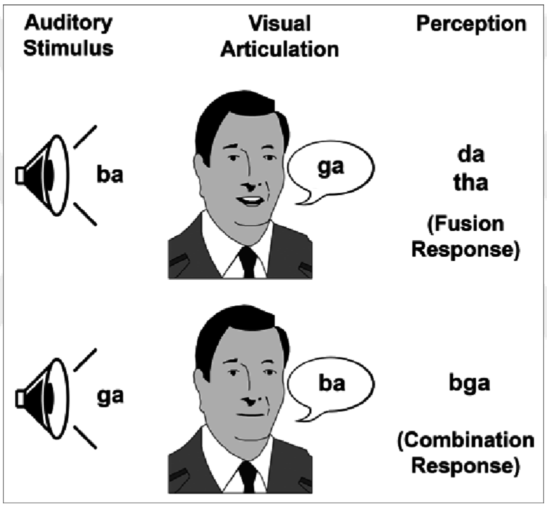


Figure 9: General concept of the McGurk Effect: Different mouth movements for auditory stimulus yield different actual sound perceptions. (From the paper by Clark, 2008. [3])

2.4.2 Experimental Evidence and Replications

The robustness of the McGurk effect has been demonstrated across numerous studies involving diverse populations and experimental conditions. Replications of the original experiment have shown that the effect persists across different languages, age groups, and even in individuals with certain sensory impairments.

For instance, a study by Burnham and Dodd (2004) [24] investigated the McGurk effect in children with and without hearing impairments. They found that while the effect was generally weaker in children with hearing impairments, it was still present, indicating that visual information plays a crucial role in speech perception even when auditory input is compromised. Other studies have explored the effect in different cultural contexts, revealing that while the strength and nature of the effect can vary, the fundamental phenomenon is universal.

2.4.3 Applications and Implications

Understanding the McGurk effect has significant practical applications, particularly in the design of audiovisual technologies, speech recognition systems, and hearing aids. By recognizing the importance of visual cues in speech perception, engineers and designers can create more effective communication devices that better support individuals with hearing impairments.

Theze et al. examined an alternate approach to stimulus design for audiovisual speech analysis using computer-generated speakers and synthetic speech [25]. They conducted an experiment in which the participants were given complete and meaningful French sentences with a mismatched audiovisual pairs based on the McGurk effect, such as "v-b" and "f-p". The difference between their stimuli and those utilized in a McGurk experiment was that they chose specific auditory and visual signals that led perception to reflect either the auditory or the visual speech token. The interactions between the differences in the visual and aural cues in a virtual reality setting are examined by Siddig et al. in [5]. Inspired by the McGurk effect, they conduct an experiment on speech perception and study the effects of audio directionality and visual quality on the degree of audiovisual integration. Van et al. investigated the McGurk effect in order to evaluate its value in advancing the knowledge of audiovisual speech perception [26].

In addition, the McGurk effect can have implications for education and teaching methods. Educators can utilize audiovisual teaching methods that leverage the integrative nature of speech perception to enhance learning outcomes. Additionally, facial identity in teaching is considered as an important factor. For instance, Walker et al. showed that facial identity and the processing of facial speech are interconnected, by observing that participants who recognize the faces are less influenced by the McGurk effect compared to those who do not [27]. Similarly, Fang et al. stated that faculty members can use students' facial expressions as feedback to refine their teaching methods and strategies, ultimately boosting the learning pace for all students in the classroom [28]. Moreover, the effect underscores the importance of considering multi-modal inputs in psychological and neuroscientific research, as it highlights the complex interplay between different sensory systems in shaping human perception.

Another way the McGurk effect can be implied as an application is incorporating the design of instructional strategies that emphasize visual articulatory feedback. For example, using video-based exercises or interactive multimedia tools that utilize visual and auditory information can aid in learning activities. As an example, Bajrami et al. stated that using suitable auditory-visual materials can enhance student-centered learning, foster student interest and engagement, and encourage greater participation, motivation, and confidence in developing communicative language skills [29].

Furthermore, the McGurk effect can highlight the role of perceptual integration and sensory adaptation in language acquisition. As an example, Hayashi et al. investigated how audiovisual speech perception through the McGurk effect is influenced when the speakers are either non-native or foreign [30]. By exploring how individuals reconcile conflicting sensory inputs, educators and researchers gain insights into cognitive processes underlying language learning and speech perception. This understanding can inform the development of personalized learning interventions tailored to individual learners' sensory preferences and abilities.

In conclusion, the implications of the McGurk effect for education and language learning underscore the benefits of integrating visual and auditory modalities to enhance comprehension, pronunciation accuracy, and overall linguistic proficiency. Embracing multimodal learning strategies can empower learners to navigate complex linguistic environments more effectively, fostering a deeper understanding and appreciation of language diversity and communication dynamics.

CHAPTER 3

METHODOLOGY

Having explained the essential concepts in the preceding chapters, this chapter outlines the proposed methodology of this study, which is organized into nine main sections. In Section 3.1, the research design is introduced. In Section 3.2, the study's objectives, research questions, associated design, instruments, and analysis tools are presented. In Section 3.3, details regarding the participants of the user study are provided. In Section 3.4, the 3D modeling of a human head to replicate the McGurk effect in a virtual environment (as previously demonstrated in [5]), including the voice recordings, mesh simplification algorithms, and the process of generating different meshes with varying resolutions, is explained. This section also explores how the mesh simplification algorithms are applied, the selection of appropriate parameter presets for each algorithm, and the essentials of setting up the virtual scene. The design of the user study is also discussed in detail. In Section 3.5, the data collection procedure and the test setup are outlined. In Section 3.6, the evaluation and comparison of the algorithms using the user study are presented, along with a results table. Finally, in Sections 3.7, 3.8, and 3.9 an overview of the assumptions, limitations, and delimitations of the study is provided, respectively.

3.1 Research Design

This study is designed to assess the cost effectiveness of three widely recognized mesh simplification algorithms detailed in Chapter 2, and incorporates quantitative metrics to rigorously evaluate and compare the effectiveness of each algorithm. Each of these algorithms is based on established theoretical frameworks and methodologies that are widely used in the field of computational geometry and computer graphics.

This study aims to make a comparative analysis between the algorithms not only in terms of computational efficiency but also in their ability to preserve visual realism during the simplification process. To achieve this, a research design based on quantitative comparative analysis is performed between three mesh simplification algorithms. The main type of quantitative research design in this study is an experimental research design, which utilizes the data gathered from a user study, in which the performance of each algorithm is quantitatively tested under varying levels of simplification. The degree of realism maintained will be evaluated using a statistical framework, allowing for direct comparison across different algorithmic implementations and simplification levels. A 3D model will serve as the core medium for this comparison, providing a virtual environment in which the impact of each simplification method on the visual fidelity of the model can be assessed. By incorporating a virtual illusion through this 3D representation, this study measures how well each algorithm balances reducing computational cost and preserving perceptual accuracy. The integration of a 3D model enables a nuanced comparison of the algorithms, while the use of statistical analysis ensures the reliability and objectivity of the findings.

3.2 Research Question

The aim of this study is to evaluate the cost effectiveness of three prominent mesh simplification algorithms, as detailed in Chapter 2. Although the implementation of these algorithms may differ across various software tools, all three are grounded in well-established methodologies: "Quadric Error Metrics," "Triangular Mesh Simplification based on Surface Angle," and "Hierarchical Face Clustering on Polygonal Meshes." To facilitate a comparative analysis, this study investigates the extent to which each algorithm preserves visual realism. This is achieved through a testing framework, which will be referred as the user study throughout this study, which enables the statistical comparison of different levels of realism. A 3D model is used to generate and sustain a virtual illusion for the purposes of this evaluation. An analysis of the responses from voluntary participants of the user study is used to assess how these different mesh simplification algorithms compare against each other in terms of maintaining the McGurk illusion.

The following research question is investigated in this study:

- Is there a significant difference between three different mesh simplification algorithms with two vertex reduction levels (35% and 70%) on the maintenance of the McGurk effect in 3D head models used for lip-sync animation?

In the Table 1 below, research question, design, instrumentation, and analysis method are summarized for this research question.

Table 1: Summary of Research Question, Design, Instrumentation, and Analysis Method

Research Question	Design	Instrumentation	Analysis
Is there a significant difference between three different mesh simplification algorithms with two vertex reduction levels (35% and 70%) on the maintenance of the McGurk effect in 3D head models used for lip-sync animation?	Quantitative comparative analysis	**Data Collection:** Comparative analysis of the scores from the user study	ANOVA

3.3 Participants

The user study involved 42 participants and participation to this user study was completely voluntary. It is important to narrow the profile of the participants using a demographic filtering, such as age and ability to see and hear clearly. Because, it was previously shown that susceptibility to the McGurk effect varies with age and development [31], as well as sensory impairment [32]. In light of this, data collected from the user study were restricted to a specific demographic group, consisting of individuals aged 18 to 35, who were only from Turkey. Since the user study involves sensory modalities of sight and hearing, participants must have no impairments in either vision or hearing. The selection criteria for participants were therefore limited to those who are aged 18 to 35, who do not experience issues with sight or hearing and who correctly answer the sound check questions. Factors such as gender, employment status, or marital status were not considered relevant criteria for inclusion in this user study. Researchers following this methodology may conduct the user study and analyze the results based on these criteria, depending on the specific focus or scope of their research. Comprehensive details about the participants are provided in Table 2.

Table 2: List of Participants

Participants	**Age**	**Impairment**
P1	18-35	None
P2	18-35	None
P3	18-35	None
P4	18-35	Visual impairment
P5	36-50	None
P6	36-50	None
P7	18-35	None
P8	36-50	None
P9	18-35	Visual and auditory impairment
P10	51-70	None
P11	18-35	None
P12	18-35	None
P13	18-35	None
P14	18-35	None
P15	18-35	None
P16	18-35	None
P17	18-35	None
P18	18-35	None
P19	18-35	None
P20	18-35	None
P21	18-35	None
P22	36-50	None
P23	18-35	None
P24	18-35	None
P25	18-35	None
P26	36-50	None
P27	18-35	None
P28	36-50	None
P29	18-35	None
P30	18-35	None
P31	18-35	None
P32	18-35	None
P33	36-50	None
P34	18-35	None
P35	36-50	None
P36	18-35	None
P37	18-35	None
P38	36-50	None
P39	18-35	None
P40	18-35	None
P41	51-70	None
P42	18-35	None

3.4 Materials

The McGurk illusion was produced, in this study, in a virtual 3D environment as it was previously studied [5] and the effect of the aforementioned simplification algorithms on the continuity of this illusion was examined. In order to create this illusion, basically, a 3D model and special rigging effects were used to create the virtual illusion in the 3D modeling and rendering software, Blender. The model was then exposed to the mesh simplification algorithms to gather different variants of the model with certain levels of vertex reductions. Since reducing the number of vertices yielded a better storage consumption, how well the illusion was preserved against volunteered participants provided guidance to the cost effectiveness of these algorithms.

In order to sufficiently produce a McGurk effect in a virtual environment, the proposed method to accomplish this can be simplified as the following procedure:

1. First, using a 3D modeling tool, design a human head mesh and apply textures and materials to it so that it becomes as convincing as possible,
2. Rig it with a proper mouth skeletal structure and create distinctly different mouth movement animations, one pair of animation for each sound, synchronized with the pronunciation of a different syllable,
3. Record and process sounds of a human who is enunciating various syllables, with a distinct recording for each syllable,
4. Setup a virtual scene, in which the head model is facing the camera directly with a clear sighting,
5. Apply the prerecorded syllable sounds as voice-overs behind each mouth animation,
6. Apply the previously mentioned mesh simplification algorithms to head model,
7. Record videos of each scene with a different algorithm-animation-sound pair.

In the end, when videos of different animations with different sounds are recorded, the necessary materials, to produce the McGurk effect, becomes apparent. In the following subsections, the details of this procedure are explained in more detail.

3.4.1 Designing, Rigging and Animating the Head Model

For the illusion to have a remarkable effect, it is imperative to design a convincing human head and set it in a convincing virtual scene. For those purposes, an adequate 3D modeling tool becomes of high importance.

The mouth animations, especially the movement of main articulators such as lips, should provide a well synchronized visual of the pronunciation of each syllable. In order to accomplish such a task, one must understand the general mechanics of phonetics, and more importantly for this study, understand if there are patterns for specific sounds. As it is explained in the book by Ladefoged [33], some sounds are produced using same mouth articulations, whereas others using different. Thus, it can be inferred that when animating the head model in this study, designing the overall movement of the mouth required making such a research in order to be effective. Indeed, with the research done by Sumpeno et al. [34], it can be seen that sounds can be mapped to distinct mouth movements.

3.4.2 Recording the Voices

Voice recordings should be obtained to match the phonetic movements of the 3D model. These recordings should be edited and fine-tuned to ensure clarity and temporal alignment with the animations. This could be done with tools such as Adobe Audition, Audacity or similar audio editing software. The selection of voice clips should be based on phonetic variations known to elicit the McGurk effect, such as the combination of different consonant and vowel sounds. As in the paper by Siddig et al. [5], utterances of particular syllables, such as /ba/ and /ma/, yield higher accuracy to observe the McGurk effect.

3.4.3 Application of the Mesh Simplification Algorithms

Since the main aim of this study is to lay out a comparative analysis of various mesh simplification algorithms, the designed model that is explained in the previous sections was used to fulfill this purpose. In light of this, the important point is to identify the key parameters for the algorithms to provide the same reductions of the number of vertices, rather than narrowing the research for specific algorithms. The concepts of key parameters or how an algorithm decimates a mesh can vary for each algorithm.

3.4.4 Key Parameter Values and Different Resolutions of the Head Model

Each algorithm has a key variable as its parameter that directly affects the number of vertices in the outputted mesh. These key variables differ from algorithm to algorithm and, to have a fair and scientific analysis, choosing the values for each parameter needs to be based on a particular set of data. In this study, the basis is the number of vertices in each output mesh, meaning that no matter which algorithm is used, the McGurk effect will be evaluated on meshes with the same number of vertices. Thus, certain levels of vertex reduction are required to make a quantitative analysis and the parameter values will then need to be calculated accordingly. An illustration about the relation between vertex counts and key parameters is shown in Figure 10.

To further elaborate on this, consider the first algorithm, the Quadric Error Metrics (QEM). Although its implementations may vary, one example is a specific implementation in Blender, in which the algorithm is used as the mechanic behind the "Decimate Geometry" tool [35]. For this implementation, the key variable is the percentage factor, by which the original mesh's vertex count will be reduced to. Similarly, for the implementation of "Limited Dissolve" [15], the key variable is the maximum allowed angle between consecutive faces, whereas for the "Un-subdivide" algorithm [16], the key variable is the number of decimation iterations.

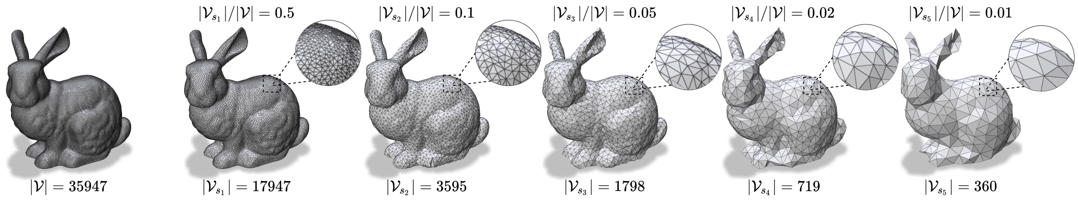


Figure 10: An example of an application of a relatively newly researched algorithm. Below value represents the total vertex count in the mesh, whilst the above value indicates the key parameter values for each resolution. Notice the different resolutions of the mesh obtained by different key parameter values. (From the paper by Potamias et al. [4])

The head model in the user study was produced with two simplified levels for each algorithm: medium and high. Each level represented a different resolution of the head mesh. For the medium level of decimation, the number of vertices were reduced by 35%, and 70% for the high level of decimation. Using the Blender implementations of the three algorithms in question, these two vertex reduction levels could be obtained for each of the algorithms by adjusting their respective parameters as follows:

- Algorithm 1 (Quadric Error Metrics)
  - Medium: Decimate Geometry ratio $\longrightarrow 0.7$
  - High: Decimate Geometry ratio $\longrightarrow 0.35$
- Algorithm 2 (Triangular Mesh Simplification based on Surface Angle)
  - Medium: Limited Dissolve max angle $\longrightarrow 11.4°$
  - High: Limited Dissolve max angle $\longrightarrow 18.3°$
- Algorithm 3 (Hierarchical Face Clustering on Polygonal Meshes)
  - Medium: Un-subdivide $\longrightarrow 1$
  - High: Un-subdivide $\longrightarrow 2$

Using each mesh simplification algorithm and the key parameter values, the mesh of the head model will be reduced to have certain numbers of vertices. Each level of simplification will play a key role for representing a key factor for the comparative analysis evaluation in the user study. Since each algorithm may provide a different approach, in this case a different key parameter, to decimate a mesh, in order to prepare a user study, a matrix consisting of each algorithm and their respective key factors should be obtained. Each level of decimation will be used in the scoring of each algorithm in the evaluation section. An example of such matrix is provided in Table 3. Algorithm 1 represents "Decimate Geometry" implementation, Algorithm 2 represents "Limited Dissolve" implementation and Algorithm 3 represents "Un-subdivide" implementation in Blender, which are implementations of "Quadric Error Metrics" [14], "Triangular Mesh Simplification based on Surface Angle" [15] and "Hierarchical Face Clustering on Polygonal Meshes" [16], respectively. Notice that, in Table 3, Algorithm 1 requires a percentage, whereas an angle value is used for Algorithm 2 and a positive integer for Algorithm 3. Using these factor values the same number of vertices are obtained at every step.

Table 3: An example table showing a matrix of different algorithms by the number of vertices for each step.

	Algorithm 1	Algorithm 2	Algorithm 3
Base Resolution (Vertex # / Factor)	10000 / %100	10000 / 0.0°	10000 / 0
Resolution 1 (Vertex # / Factor)	7000 / %70	7000 / 11.4°	7000 / 1
Resolution 2 (Vertex # / Factor)	3500 / %35	3500 / 18.3°	3500 / 2

As a different visual example, head models that includes different number of vertices as 5815, 2927, 1476, 894 and 602, are provided in Figures 11, 12, 13, 14 and 15 respectively.

Figure 11: The base head model with number of vertices of 5815.

Figure 12: A head model with number of vertices of 2927, reduced by mesh simplification.

Figure 13: A head model with number of vertices of 1476, reduced by mesh simplification.

Figure 14: A head model with number of vertices of 894, reduced by mesh simplification.

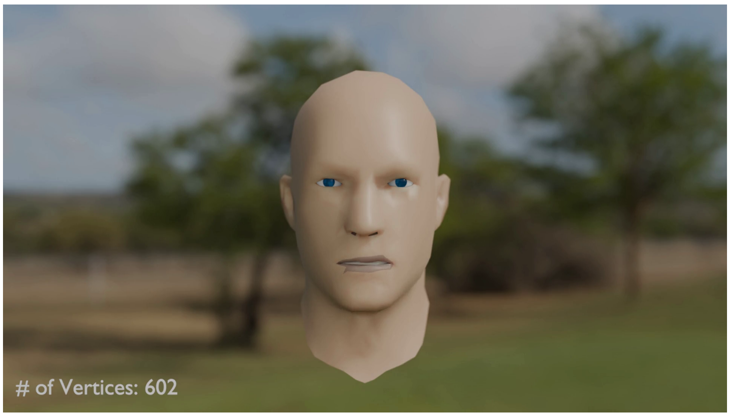


Figure 15: A head model with number of vertices of 602, reduced by mesh simplification.

3.4.5 Virtual Scene Setup Essentials

In order to compile the major assets that are gathered together and create a convincing virtual scene, in addition to the head model, other essential components should be provided; most importantly, the camera and the lighting. Since the illusion is simulated in a virtual environment, proper camera setup and proper lighting becomes crucial: it should be as if the person is standing in front of the model. For the McGurk effect to be observed most prominently, as in the research by Siddig et al. [5], the head should be placed in a near field (even in virtual conditions) and facing front towards the camera. Examples of the near (2 m.) and far (4 m.) fields are shown in Figure 16, which provides an example of a virtual setup. Additionally, in the aforementioned paper, it is shown that the virtual locations of the sound adversely alters the amount of illusion. For this reason, to maximize the impact of the illusion and since the illusion depends on the audio-visual perception, the sound should only come from the front side. Lastly, the background should have a nice contrast and the environment should be well lit, in order to allow the head to be clearly seen by the observer.

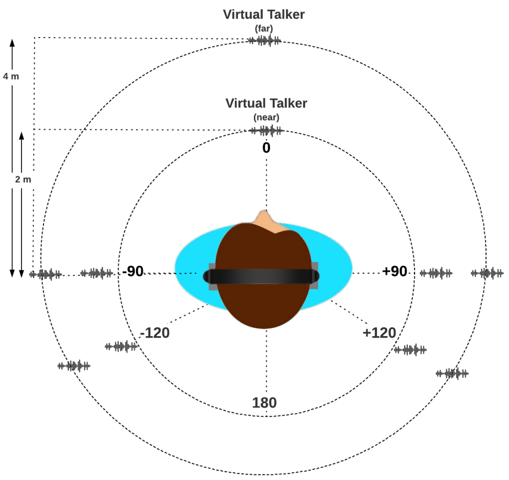


Figure 16: An example virtual setup of an observer seen in top-down view. The listener in the center shall be the virtual camera. (From the paper by Siddig et al. [5].)

3.4.6 The Videos and Questions of the User Study

To provide an answer to the research question proposed in this study, a specifically formulated user study can be conducted to assess a comparison relation between any collection of mesh simplification algorithms. The way to achieve this is to have the user study being composed of a sequence of multiple choice questions, with each question presenting the participant with an audible video associated with that question. The videos should be prepared using the materials that are mentioned in the previous sections; a head model with articulated mouth animations, sound recordings of the pronunciations of designated syllables and a virtual 3D scene, in which the head model resides with a camera and lighting setup. Each prepared video should be placed in the aforementioned multiple choice questions, one video per question, and at each question the answers should include the textual representation of the correct sound that is used in the making of the associated video, as well as the textual representations of some other wrong sounds. It must be clearly indicated that, the goal of each and every question in the test is to check, after playing the video, if the participant hears and chooses the correct sound as the answer.

In this study, a preliminary work was conducted to better understand the processes involved in preparing the videos to be shown to participants in the user study. As part of the preliminary study, audio recordings of three different syllables were made: ’ba’, ’ma’, and ’va’, similar to those used by Siddig et al. in [5].

The audio recordings were made by the researcher, who would also conduct the user study, ensuring that each recording maintained consistent pitch, monotony, and normalized intensity. In parallel, a three-dimensional human face model was created and two distinct mouth-lip movement animations were applied to it. For each animation, one short silent video was recorded. Additionally, short silent videos were made using the same two animations but with the model’s vertices decimated using three different mesh simplification algorithms and two simplification levels (35% and 70%). A total of 12 videos were produced (3 algorithms × 2 simplification levels × 2 animations). The simplification levels were chosen as follows: beyond 70% simplification, the mesh becomes distorted and unrecognizable, so 70% simplification was considered the high level. The medium simplification level, set at 35%, was half of this value.

Subsequently, voice-over recordings were added to each of the 14 silent videos (2 original animations + 12 simplified versions). For each silent video, the three-syllable audio recording was synchronized with the model’s mouth movements to create the illusion of speech, and the voice-over was re-recorded. As a result, 14 voiced videos were created, totaling 42 videos for the three syllables. Additionally, 3 videos were made in which the syllables were voiced but without the 3D face model, leaving only the audio and a blank background. In total, 45 videos were produced.

The number of videos, and correspondingly the number of questions in the user study depends on the materials provided by the preliminary work. In detail, the materials can be formulated as the following:

- Let $S = \{s_1, s_2, ..., s_x\}$ be the collection of recorded syllable sounds,
- Let $M = \{s_1m_1, s_1m_2, s_2m_1, s_2m_2, ..., s_xm_1, s_xm_2\}$ be the collection of distinctly different articulated mouth animation pairs for each sound – there may be duplicate animations, which will not interfere with or affect the results of the study,
- Let $A = \{a_1, a_2, ..., a_y\}$ be the collection of the mesh simplification algorithms that are compared,
- For each algorithm i, let $A_i = \{a_if_1, a_if_2, ..., a_if_k\}$ be the collection of different resolutions of the mesh.

Given these preliminary materials, the questions in the user study were prepared according to the following procedure:

1. Each participant should be checked for the quality of their audibility, meaning, whether they can hear correctly. Thus, using the collection S, prepare x number of videos for each sound without any 3D elements. These videos should be presented as the first sound check (or prime) questions, namely:
$$Q' = \{s_1, s_2, ..., s_x\}$$

2. Moving on after the prime questions, combinations of only the sound-animation pairs from the collections S and M without any mesh simplification, should be produced. The videos of this production can be designated as the zero questions because no mesh simplification is applied:

$$Q_0 = \left\{ \begin{array}{l} (s_1, s_1 m_1), (s_1, s_1 m_2) \\ (s_2, s_2 m_1), (s_2, s_2 m_2) \\ \dots \\ (s_x, s_x m_1), (s_x, s_x m_2) \end{array} \right\}$$

3. Finally the really important videos are produced; which are the combination of both x number of sounds, a pair of animations for each sound and y number of mesh simplification algorithms and k number of resolutions for each algorithm. For each algorithm i, the question set becomes:

$$Q_i = \left\{ \begin{array}{l} (s_1, s_1 m_1, a_i, f_1), (s_1, s_1 m_1, a_i, f_2) \dots (s_1, s_1 m_1, a_i, f_k), \\ (s_1, s_1 m_2, a_i, f_1), (s_1, s_1 m_2, a_i, f_2) \dots (s_1, s_1 m_2, a_i, f_k), \\ \\ \dots \\ (s_2, s_2 m_1, a_i, f_1), (s_2, s_2 m_1, a_i, f_2) \dots (s_2, s_2 m_1, a_i, f_k), \\ (s_2, s_2 m_2, a_i, f_1), (s_2, s_2 m_2, a_i, f_2) \dots (s_2, s_2 m_2, a_i, f_k), \\ \\ \dots \\ (s_x, s_x m_1, a_i, f_1), (s_x, s_x m_1, a_i, f_2) \dots (s_x, s_x m_1, a_i, f_k), \\ (s_x, s_x m_2, a_i, f_1), (s_x, s_x m_2, a_i, f_2) \dots (s_x, s_x m_2, a_i, f_k), \end{array} \right\}$$

4. This yields to a total number of videos $V = x + (2x) + (2x * y * k)$. It should be noted that, in order to eliminate any bias or guessing by any participant, the videos in the questions should be presented in a random order, without any dependency on the order of their creations. The designer of the study can store the information about which video is created with which materials, without disclaiming it to the participants.

All of the videos that are prepared for this user study is provided in Table 4 below:

Table 4: All 42 videos prepared for the user study, sorted in four categories; no simplification and simplification using algorithms 1, 2 and 3. Within each category, videos are sub-sorted using different syllables and mouth animations.

Base Videos	Algorithm 1	Algorithm 2	Algorithm 3
1. Ba_anim1_base	7. Ba_anim1_alg1_med	19. Ba_anim1_alg2_med	31. Ba_anim1_alg3_med
2. Ba_anim2_base	8. Ba_anim1_alg1_high	20. Ba_anim1_alg2_high	32. Ba_anim1_alg3_high
3. Ma_anim1_base	9. Ba_anim2_alg1_med	21. Ba_anim2_alg2_med	33. Ba_anim2_alg3_med
4. Ma_anim2_base	10. Ba_anim2_alg1_high	22. Ba_anim2_alg2_high	34. Ba_anim2_alg3_high
5. Va_anim1_base	11. Ma_anim1_alg1_med	23. Ma_anim1_alg2_med	35. Ma_anim1_alg3_med
6. Va_anim2_base	12. Ma_anim1_alg1_high	24. Ma_anim1_alg2_high	36. Ma_anim1_alg3_high
	13. Ma_anim2_alg1_med	25. Ma_anim2_alg2_med	37. Ma_anim2_alg3_med
	14. Ma_anim2_alg1_high	26. Ma_anim2_alg2_high	38. Ma_anim2_alg3_high
	15. Va_anim1_alg1_med	27. Va_anim1_alg2_med	39. Va_anim1_alg3_med
	16. Va_anim1_alg1_high	28. Va_anim1_alg2_high	40. Va_anim1_alg3_high
	17. Va_anim2_alg1_med	29. Va_anim2_alg2_med	41. Va_anim2_alg3_med
	18. Va_anim2_alg1_high	30. Va_anim2_alg2_high	42. Va_anim2_alg3_high

A list of numbers ranging from 1 to 42, in a random order can be used for correlating videos indices and question numbers. The list, which can be named V_i, that was used in the user study is as follows:

$$V_i = \left\{ \begin{array}{c} 13, 33, 21, 29, 15, 4, 1, 17, 25, 2, 3, 14, 11, 8, 19, 20, 23, 26, 30, 24, 36, \\ 32, 38, 12, 42, 7, 31, 41, 9, 10, 35, 28, 39, 16, 27, 6, 18, 40, 22, 5, 37, 34 \end{array} \right\}$$

The way this list was used is per the following: The position of an element in V_i indicates its question number after the control questions. For example; the 2nd element, which is 33, in V_i means that the video 33 (Ba_anim2_alg3_med) is the 2nd question after the first three control questions; that is Question 5 from the start. In order to evaluate the user study, a map is needed to correlate each question's video with the video table, referenced in Table 4. Using the random generated video table index list V_i, the map can be obtained as per the following:

Table 5: The question map that is used in the evaluation process of the user study. Notice the video table index column is actually the list V_i.

Video Table Index	Question No (After the Control Q.s)
13 (Ma_anim2_alg1_med)	1
33 (Ba_anim2_alg3_med)	2
21 (Ba_anim2_alg2_med)	3
29 (Va_anim2_alg2_med)	4
...	...
5 (Va_anim1_base)	40
37 (Ma_anim2_alg3_med)	41
34 (Ba_anim2_alg3_high)	42

3.5 Procedure

Since the critical evaluation process focuses on observing the McGurk effect, a reliable method for detecting the illusion should be established. Before delving into further details, it is important to note, as highlighted in [5], that the strength of the McGurk effect varies in accuracy depending on the syllable. As a result, the comparison of mesh simplification algorithms should be categorized by individual sound. Using the variables defined in the previous section, this will result in x syllable categories, with each category representing a comparative analysis based on that particular syllable.

A multiple-choice questionnaire was administered to gather data for the user study. This questionnaire consisted of 45 questions, each accompanied by a short video (1-2 seconds in length) relevant to the question. Participants were asked to select one of several possible answers based on the content of the video. The videos were embedded within the questionnaire and could be viewed as many times as needed. To ensure valid data collection, participants were required to answer all questions in full. The first three questions served as control items, designed to verify that participants provided appropriate responses throughout the remainder of the questionnaire, and participants were informed of this. The entire questionnaire took approximately 7-10 minutes to complete, as anticipated. The full questionnaire is provided in Appendix C.

In the user study involving the 42 participants, as previously mentioned, each participant is first required to complete the demographic questions. Only the ones who pass these demographic filtering questions were considered for the analysis phase. Then, for each 3 syllables ('va', 'ma' and 'ba'), there were respective control sound questions in the test. These 3 questions presented no head model but only the syllables' respective sounds. Participants were then asked which sounds they heard in those questions. These sound check questions were critical; meaning that, if a participant fails the sound check by incorrectly identifying the sound source for the video, all subsequent questions related to the failed syllable will be excluded from analysis.

Moving on to the head model questions, there were 12 questions per syllable category, as explained in Section 3.4.6. In these questions, participants were also asked which sound they heard and were required to answer accordingly. Points accumulated for each mesh simplification algorithm for each syllable category. The way points accumulate is as following: For each participant the McGurk effect should be maintained, even with head models reduced with mesh simplification algorithms' medium (35%) and high reduction (70%) levels. Maintaining the illusion at both levels of reduction grants each algorithm a maximum of 1 point for each participant who was affected by the illusion. For a participant to be considered as affected, that participant should answer the control sound question correctly for that syllable, give different answers for each syllable-animation combination on non-simplified versions of the head model and give different answers for each syllable-animation combination on medium and high reduction levels. Being affected at medium and high reduction levels grant the algorithm in question specific points, based on the difficulty/bias weighting between medium and high factors. Using the weighting adjustments depending on the level of bias, each algorithm will have varying scores based on different bias levels for that syllable, with a maximum of 1 point when both at medium and high factors the McGurk is observed. This will repeat for each syllable category and for each participant.

As a summary, in the first three questions of the questionnaire, participants were presented with control questions, where they were asked to identify the sound they heard from the voice recordings of the syllables ’ba’, ’ma’, and ’va’, without any three-dimensional human face being shown. In the subsequent 42 questions, participants watched model animation videos with voice-over, as previously described in Table 5, in a random order. They were then asked to identify the sound they heard in the same manner. Overall, participants provided responses about the sounds they heard for all 45 videos presented during the questionnaire.

3.6 Data Analysis

In this study, the main aim is to make a quantitative comparative analysis between three different mesh simplification algorithms. For this purpose, a scoring system and one way Analysis of Variance (ANOVA) test were used. The ANOVA test compares the means of the success scores across different algorithms to see if any significant differences exist. If, per each syllable category, the resulting P-value is less than 0.05, than it can be safely assumed that the results are not based on chance and indeed comparable. As a result of the user study, the answers given by the participants were examined one by one and statistical data was obtained with a scoring system to determine whether the McGurk illusion really occurred and if so, the performances of the model simplification algorithms was examined. The performance comparisons that emerged as a result of the examinations were categorized separately on the basis of each syllable (’ba’, ’ma’ and ’va’) voice recording. A two-phase analysis was performed in each category of the three audio recordings. In the first phase, it was firstly examined whether the McGurk effect has occurred in the participant for that syllable. If the participant gave different answers to the 2 videos where this audio recording was used, the model was not simplified and different animations were played, the effect was considered to have occurred. If it had not occurred, the answers to the questions in the videos containing the audio recordings of that syllable were not examined and the data of this participant were not included in the analysis in the category of that syllable. If an illusion was achieved, the second phase was passed and the results of each algorithm were examined among the videos containing the audio recordings of this syllable.

3.6.1 Evaluation for Each Syllable

For each category, prior to examining the McGurk effect in videos with mesh simplification applied, the illusion must first be observed in videos (and their corresponding questions) that have no simplification applied. These videos are described in the question preparation procedure, Step 2. To verify the presence of the McGurk effect in these types of questions, the following approach was used: if the sound in a question with the combination $(s_i m_1)$ is perceived differently from the sound in a question with the combination $(s_i m_2)$, it can be concluded that the McGurk effect is present for that participant in the i-th syllable, with no mesh decimation applied. Based on this, the same verification process will be repeated for each mesh simplification algorithm applied to the corresponding syllable-animation-algorithm combinations.

To collect data that can be analyzed, a clear method of quantification for each syllable-animation-algorithm combination should be established. As outlined before, the user study involves questions with different resolutions of the head mesh, and each resolution factor should contribute to the overall score in a way that reflects its difficulty level. The key challenge here is determining how to score each level of difficulty associated with these resolutions. Since there is no precise benchmark for comparing these difficulty levels, a biased scale can be employed, ranging from equal scores for each level to a higher bias for more difficult levels. This varying bias can serve as a weight adjustment, following a statistical weighting technique by Kalton et al. [36], to appropriately score each factor and highlight the varying degrees of difficulty in displaying the McGurk effect. The use of this bias range will enable the creation of a graphical representation for the comparative analysis of each algorithm.

In the proposed methodology, the effectiveness of the McGurk effect for each algorithm was quantified on a scale from 0 (indicating no effectiveness) to 1 (indicating full effectiveness). Let k represent the total number of resolutions. For resolutions numbered 1 through k, let $\{f_1, f_2, ..., f_k\}$ denote the set of resolutions for the head mesh. Let $S(r, b)$ represent the score for the r-th resolution and the bias $b \in \mathbb{N}^+$, which corresponds to a specific factor and algorithm. Accordingly, the score for a resolution that successfully induces the McGurk illusion can be calculated as follows:

$$S(r, b) = (b^{r-1})/(\sum_{n=0}^{k-1} b^n)$$

By applying this formula and adjusting the bias starting from 1, a scoring matrix can be generated for all the algorithms under analysis. The scores for each algorithm in the set $a_1, a_2, ..., a_y$ are denoted as $X(a_q)$, which represents an array of scalar values for the q-th algorithm, calculated using varying bias levels. These scalar values are derived from the data collected from each participant. Using the resolution r, the total number of resolutions r_{max}, and the number of participants N_i analyzed for the i-th syllable, the following formula, which provides an array of scores, can be applied:

$$X(a_q, b) = \{b \in \mathbb{N}^+ | (\sum_{r=1}^{r_{max}} S(r, b))/N_i : 0 \leq \sum_{r=1}^{r_{max}} S(r, b) \leq 1\}$$

This array provides different quantification representations for each algorithm, taking into account varying difficulty levels associated with different mesh decimation resolutions. The resulting data will highlight distinct trends in the effectiveness of the McGurk effect for each algorithm, if present. Significant differences will allow for clear performance comparisons between the algorithms for each syllable.

Table 6: An template score result table of the user study for a single syllable, using the algorithm as shown in Algorithm 1, showing the final scores for the three compared algorithms.

	Bias 1	Bias 2	Bias 3	...
Alg. 1	$X(a_1, 1)$	$X(a_1, 2)$	$X(a_1, 3)$	...
Alg. 2	$X(a_2, 1)$	$X(a_2, 2)$	$X(a_2, 3)$	...
Alg. 3	$X(a_3, 1)$	$X(a_3, 2)$	$X(a_3, 3)$	...

The two axes of the score tables are i) the algorithms used in simplifying the base videos and ii) the (difficulty) bias ratios for scoring each algorithm. In order to obtain a score for each cell of this table, corresponding accumulated points for that algorithm using that bias ratio must be divided by the number of participants who were affected by the McGurk illusion for that syllable. To accomplish this, the data of all the participants (who at least passed the demographic filtering) are iterated one by one, and are submitted to a series of checks. Observing the results of these checks and using the current bias level's scoring system, each algorithm's score is accumulated. The following method, which is described in a pseudocode, was invoked for each participant and each syllable, for each bias levels ranging from 1 to 10:

Input: The syllable in text form, **syllable**
Input: The syllable's control sound question no, **csNo**
Input: Control sound question answers, **csAnswers**
Input: Rest of the question answers, **qAnswers**
Input: The ‘video index‘-‘question no‘ map, **qMap**
Input: The bias level (e.g. 2, which means 1:2), **biasLevel**
Output: The number of participants who "heard" the syllable's control question, **correctHearingCount**
Output: The number of participants who were affected by the McGurk effect for this syllable, **affectedCount**
Output: Algorithm 1's accumulated points for this syllable and at this bias level, **alg1AccPoints**
Output: Algorithm 2's accumulated points for this syllable and at this bias level, **alg2AccPoints**
Output: Algorithm 3's accumulated points for this syllable and at this bias level, **alg3AccPoints**
Output: Algorithm 1's final score for this syllable and at this bias level, **alg1FinalScore**
Output: Algorithm 2's final score for this syllable and at this bias level, **alg2FinalScore**
Output: Algorithm 3's final score for this syllable and at this bias level, **alg3FinalScore**

```
participantCount ← length(csAnswers)
medPt ← 1/(1 + biasLevel)
hiPt ← 1−medPt
result ← {} // Initialize result object
for p ← 1 to participantCount do
    curParCSAnswers ← csAnswers[p]
    curParQAnswers ← qAnswers[p]
    csAns ← curParCSAnswers[csNo]
    csCorrect ← (csAns = syllable)
    if csCorrect then
        increment result.correctHearingCount
    end
    baseEffectOK ← csCorrect (curParQAnswers[qMap[syllable_anim1_base]] ≠
      curParQAnswers[qMap[syllable_anim2_base]])
    if !baseEffectOK then
        continue // Skip to the next participant
    end
    increment result.affectedCount
    // Algorithm 1
    alg1MedAffecting ← (curParQAnswers[qMap[syllable_anim1_alg1_med]] ≠
      curParQAnswers[qMap[syllable_anim2_alg1_med]])
    if alg1MedAffecting then
        result.alg1AccPoints ← result.alg1AccPoints + medPt
    end
    alg1HiAffecting ← (curParQAnswers[qMap[syllable_anim1_alg1_hi]] ≠
      curParQAnswers[qMap[syllable_anim2_alg1_hi]])
    if alg1HiAffecting then
        result.alg1AccPoints ← result.alg1AccPoints + hiPt
    end
    // Algorithm 2
    alg2MedAffecting ← (curParQAnswers[qMap[syllable_anim1_alg2_med]] ≠
      curParQAnswers[qMap[syllable_anim2_alg2_med]])
    if alg2MedAffecting then
        result.alg2AccPoints ← result.alg2AccPoints + medPt
    end
    alg2HiAffecting ← (curParQAnswers[qMap[syllable_anim1_alg2_hi]] ≠
      curParQAnswers[qMap[syllable_anim2_alg2_hi]])
    if alg2HiAffecting then
        result.alg2AccPoints ← result.alg2AccPoints + hiPt
    end
    // Algorithm 3
    alg3MedAffecting ← (curParQAnswers[qMap[syllable_anim1_alg3_med]] ≠
      curParQAnswers[qMap[syllable_anim2_alg3_med]])
    if alg3MedAffecting then
        result.alg3AccPoints ← result.alg3AccPoints + medPt
    end
    alg3HiAffecting ← (curParQAnswers[qMap[syllable_anim1_alg3_hi]] ≠
      curParQAnswers[qMap[syllable_anim2_alg3_hi]])
    if alg3HiAffecting then
        result.alg3AccPoints ← result.alg3AccPoints + hiPt
    end
end
result.alg1FinalScore ← result.alg1AccPoints / result.affectedCount
result.alg2FinalScore ← result.alg2AccPoints / result.affectedCount
result.alg3FinalScore ← result.alg3AccPoints / result.affectedCount
return result
```

Algorithm 1: The pseudocode of the method that is used to calculate the scores of a syllable, using the data from the user study.

3.6.2 Making a Comparative Analysis Using ANOVA

It is crucial to evaluate whether the results show significant differences, as this would determine whether the scores between algorithms can support a meaningful comparative analysis. To assess this, an ANOVA test can be performed for each syllable category. The ANOVA test, first introduced by Ronald A. Fisher in his foundational work "Statistical Methods for Research Workers" [37], is a statistical technique used to assess whether there are significant differences between the means of three or more independent groups. In the context of this study, which compares the cost-effectiveness of various 3D mesh simplification algorithms using the McGurk effect, the aim of the ANOVA test is to determine if there are statistically significant differences in the effectiveness of the different algorithms in maintaining the McGurk effect. More specifically, the ANOVA test will inform decisions about which algorithms are the most cost-effective and efficient for each application, thus informing future research and practical applications in 3D modeling and mesh simplification. In brief, the ANOVA test can be conducted as follows: The data for each algorithm can be organized into a matrix, where each row corresponds to a specific algorithm, and each column represents a participant's response under different conditions (such as varying biases). From this data matrix, the ANOVA test will produce two key values: the F-value and the P-value. The F-value quantifies the ratio of variance between the group means to the variance within the groups, with a higher F-value indicating a greater difference between the group means relative to the internal variance. The P-value, on the other hand, represents the probability that the observed differences between the group means occurred by chance. A low P-value (usually less than 0.05) suggests that the differences are statistically significant. For instance, a very low P-value would indicate that the differences between the algorithms' means are highly significant and unlikely to be due to random chance.

3.6.3 Ethical Considerations

Participation in the user study was conducted through Google Forms and was entirely voluntary. Participants were initially contacted via email, and those who expressed interest and agreed to participate by accessing the survey link were presented with an informed consent form at the outset, as detailed in Appendix B. Each participant signed the consent form, which was approved by the university's Human Subjects Ethics Committee, confirming their voluntary involvement in the user study. Only those who provided their consent were allowed to continue with the questionnaire.

Prior to their participation, all individuals were fully informed about the objectives and procedures of the study and provided explicit consent. The research was conducted in accordance with established ethical standards to safeguard participant privacy and confidentiality. No personally identifiable information was collected, and all data were anonymized to protect participant identity. Ethical clearance of this user study is provided in Appendix A.

3.7 Assumptions

In the scope of this study, the following can be assumed:

- Both the consent form and the questions in the user study are adequate for each participant to clearly understand and answer.

- Each participant contributes to the research voluntarily and without any negative intent that would purposefully hinder the research.
- The participants would answer honestly and without bias from any external source.
- The listening devices used by each participant relays sound output that is sufficiently good for the purposes of this study.
- The data from the user study are to be correctly gathered and analyzed.
- According to a research, published in 2024 by Magnotti et. al. [38], there might be lasting changes to auditory perception when a person is exposed to the McGurk effect repeatedly. Although this can be further studied, this research assumes that the participants are not exposed too much to the illusion to be induced with such a negative effect.

3.8 Limitations

The limitations of this study can be described as the following:

- The data gathered from the user study are limited to a certain demographic scope; from people within ages of 18 to 35 and who have problem with neither in their eyesight nor their hearing.
- The user study was conducted in the spring semester of 2023-2024, over a course of a few weeks only.
- The videos in the user study were all uploaded to YouTube™; and as such, it is imperative to note that they were all subject to the website's own video compression which could slightly downgrade the quality the videos. However, it has been shown that, in the context of experiencing the McGurk effect, even with a degradation in visual and auditory stimuli, the effect still persists [39].
- The formality of each user study question requires the use of a good set of listening materials, such as headphones or speakers, and visual displays, such as computer monitors or tablets. The quality of these materials may vary for each participant.
- The participants are selected from the country of Turkey only.

3.9 Delimitations

The delimitations of this study can be identified as the following:

- Since the user study is limited to a certain demographic scope, the sample size of the user study is limited.
- The analysis made in the study is only possible with the data gathered from the user study.

CHAPTER 4

RESULTS

In this study, three different, but also, prominent mesh simplification algorithms were selected and the main purpose of this study is to compare the cost efficiency between these algorithms using the approach that is mentioned in Chapter 3, Methodology. Although the implementations of these algorithms vary by different software tools, the three selected algorithms are, nevertheless, based on studies; "Quadric Error Metrics" [14], "Triangular Mesh Simplification based on Surface Angle" [15] and "Hierarchical Face Clustering on Polygonal Meshes" [16]. To incorporate a way of comparison for this study, it is imperative to test how the visual realism is maintained as these algorithms are utilized in a quantified testing scenario in which different levels of realism can be compared statistically. To implement such a comparison, a virtual illusion is created and maintained via a 3D model.

4.1 Survey Results and Analysis

In this section, the analysis of the user study, conducted to evaluate the effectiveness of various 3D mesh simplification algorithms in sustaining the McGurk effect, is presented. As mentioned in Section 3.8 of Chapter 3, in order to narrow the scope of the user study, an age restriction of 18-35 years was set. Additionally, this user study was conducted on only with people who are from Turkey. The study involved 42 participants, with results from 29 participants who passed a preliminary test being analyzed, as mentioned in Chapter 3. The analysis focused on three specific syllables: 'va', 'ma', and 'ba'. For each syllable, three algorithms were utilized with two effectiveness factors: medium and high. The following subsections provide a detailed analysis of the results for each syllable. Participants that answered the control sound check questions correctly and answered differently to the questions that had no mesh simplification algorithms for each syllable, showing the presence of the McGurk effect, was considered in the analysis. That group of participants will be referred as "sample space" in the following.

Algorithm 1 represents "Decimate Geometry" implementation, Algorithm 2 represents "Limited Dissolve" implementation and Algorithm 3 represents "Un-subdivide" implementation in Blender, which are implementations of "Quadric Error Metrics" [14], "Triangular Mesh Simplification based on Surface Angle" [15] and "Hierarchical Face Clustering on Polygonal Meshes" [16], respectively. Additionally, the bias ratio in the tables that can be found in the following subsections represents the ratio of the score weights of the lowest factor by the highest factor.

As per the explanation in the Methodology chapter, a one-way ANOVA test was also conducted to determine if there are statistically significant differences between each algorithm's performance for each syllable.

4.1.1 Syllable 'va'

The sample space for the syllable 'va' was 4. The results for the syllable 'va' are presented in Table 8. Each row represents an algorithm, and each column represents a different bias, with the fractions indicating the distribution of points between medium and high effectiveness. The graphical representation is provided in Figure 17.

Table 7: Results of the user study for Syllable *'va'*, using the algorithm as shown in Algorithm 1, showing the accumulated points for each of the three algorithms compared.

	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
Alg. 1	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000
Alg. 2	2.500	2.000	1.750	1.600	1.500	1.429	1.375	1.333	1.300	1.273
Alg. 3	3.500	3.667	3.750	3.800	3.833	3.857	3.875	3.889	3.900	3.909

Table 8: Results of the user study for Syllable *'va'*, showing the scores for each of the three algorithms compared. Each cell value is calculated using the respective cell value in Table 7 divided by 4 (which is the number of participants under the McGurk effect for this syllable).

	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
Alg. 1	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750
Alg. 2	0.625	0.500	0.438	0.400	0.375	0.357	0.344	0.333	0.325	0.318
Alg. 3	0.875	0.917	0.938	0.950	0.958	0.964	0.969	0.972	0.975	0.977

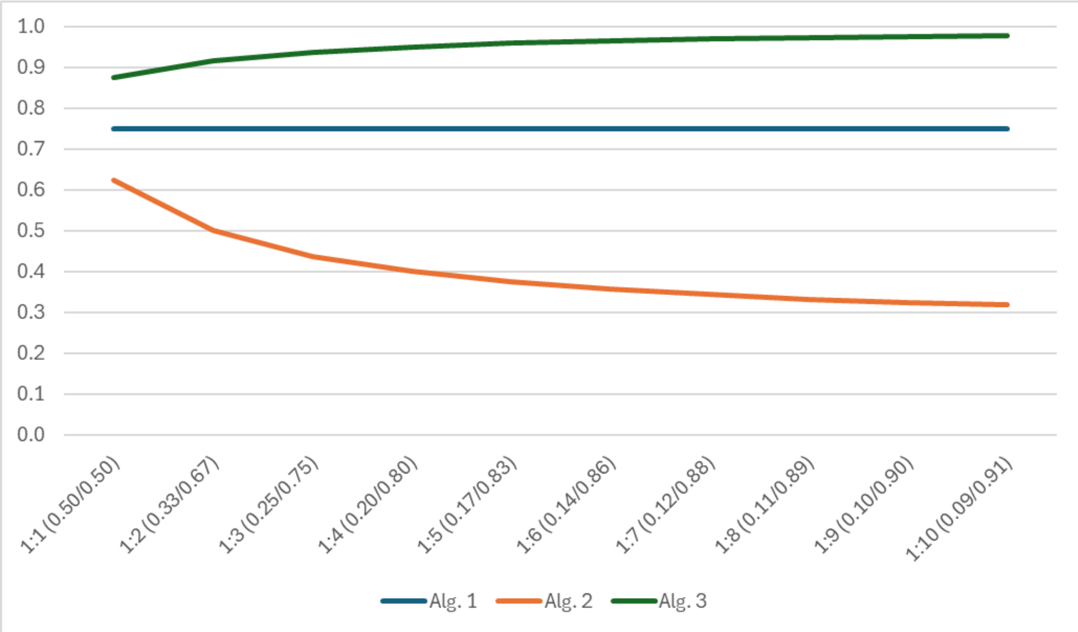


Figure 17: Graphical representation of Table 8.

From the table, it is evident that Algorithm 1 maintains a constant success score of 0.750 across all biases. Algorithm 2's success score decreases as the bias becomes more skewed towards high effectiveness, indicating a lower ability to sustain the illusion under these conditions. Conversely, Algorithm 3's success score increases as the bias becomes more skewed towards high effectiveness, demonstrating superior performance in sustaining the illusion.

Also, using the basic of the ANOVA test, the results for the syllable 'va' (based on the responses from 4 participants) are as follows:

$$F\text{-}value : 221.41$$

$$P\text{-}value : 1.79 * 10^{-17}$$

These results indicate a statistically significant difference between the means of the three algorithms. The extremely low p-value suggests that the differences observed are highly unlikely to be due to chance.

4.1.2 Syllable 'ma'

The sample space for the syllable 'ma' was 2. The reason the sample space for the syllable 'ma' is low is because most participants correctly heard the sound whatever the mouth articulation is, thus, preventing the McGurk effect to occur. The results for the syllable 'ma' are shown in Table 10. Similar to the previous table, rows represent algorithms, and columns represent different biases. The graphical representation is provided in Figure 18.

Table 9: Results of the user study for Syllable *'ma'*, using the algorithm as shown in Algorithm 1, showing the accumulated points for each of the three algorithms compared.

	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
Alg. 1	1.000	0.667	0.500	0.400	0.333	0.286	0.250	0.222	0.200	0.182
Alg. 2	0.500	0.667	0.750	0.800	0.833	0.857	0.875	0.889	0.900	0.909
Alg. 3	0.500	0.667	0.750	0.800	0.833	0.857	0.875	0.889	0.900	0.909

Table 10: Results of the user study for Syllable *'ma'*, showing the scores for each of the three algorithms compared. Each cell value is calculated using the respective cell value in Table 9 divided by 2 (which is the number of participants under the McGurk effect for this syllable).

	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
Alg. 1	0.500	0.333	0.250	0.200	0.167	0.143	0.125	0.111	0.100	0.091
Alg. 2	0.250	0.333	0.375	0.400	0.417	0.429	0.438	0.444	0.450	0.455
Alg. 3	0.250	0.333	0.375	0.400	0.417	0.429	0.438	0.444	0.450	0.455

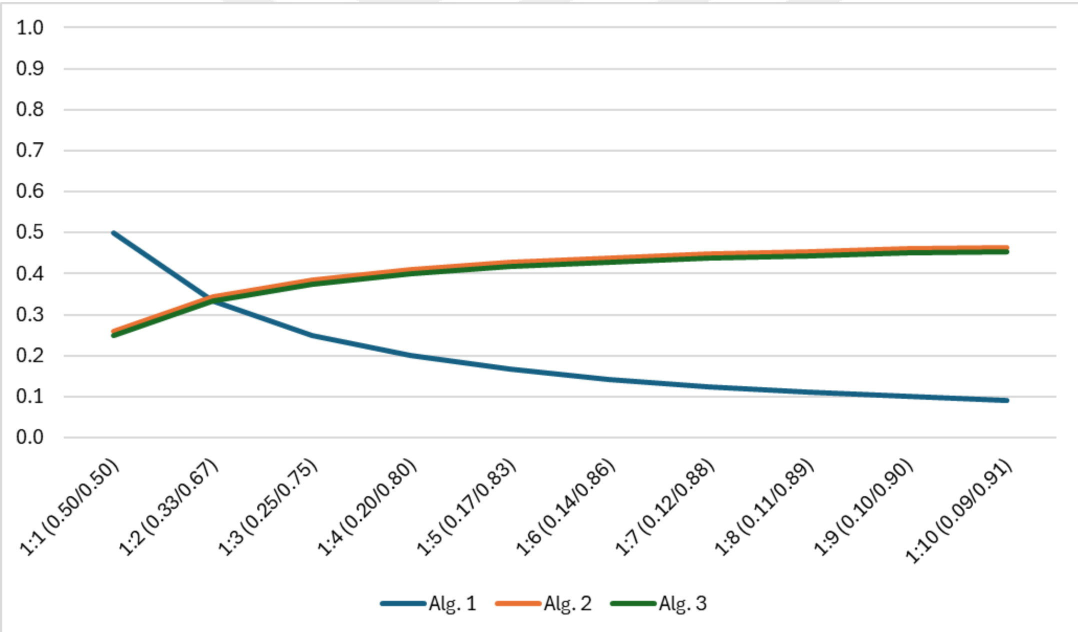


Figure 18: Graphical representation of Table 10.

The data indicates that Algorithm 1's success score decreases significantly as the bias becomes more skewed towards high effectiveness. In contrast, Algorithms 2 and 3 show a relatively stable performance, with their success scores slightly increasing as the bias shifts. This suggests that Algorithms 2 and 3 are more robust in sustaining the illusion for the syllable 'ma' under varying conditions.

The ANOVA test results for the syllable 'ma' (based on the responses from 2 participants) are as follows:

$$F\text{-}value : 15.53$$

$$P\text{-}value : 3.25 * 10^{-5}$$

These results indicate a statistically significant difference between the means of the three algorithms. The very low p-value suggests that the differences observed are highly unlikely to be due to chance.

4.1.3 Syllable 'ba'

The sample space for the syllable 'ba' was 10. The results for the syllable 'ba' are detailed in Table 12. The graphical representation is provided in Figure 19.

Table 11: Results of the user study for Syllable *'ba'*, using the algorithm as shown in Algorithm 1, showing the accumulated points for each of the three algorithms compared.

	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
Alg. 1	6.000	6.333	6.500	6.600	6.667	6.714	6.750	6.778	6.800	6.818
Alg. 2	5.000	5.667	6.000	6.200	6.333	6.429	6.500	6.556	6.600	6.636
Alg. 3	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000

Table 12: Results of the user study for Syllable *'ba'*, showing the scores for each of the three algorithms compared. Each cell value is calculated using the respective cell value in Table 11 divided by 10 (which is the number of participants under the McGurk effect for this syllable).

	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
Alg. 1	0.600	0.633	0.650	0.660	0.667	0.671	0.675	0.678	0.680	0.682
Alg. 2	0.500	0.567	0.600	0.620	0.633	0.643	0.650	0.656	0.660	0.664
Alg. 3	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600

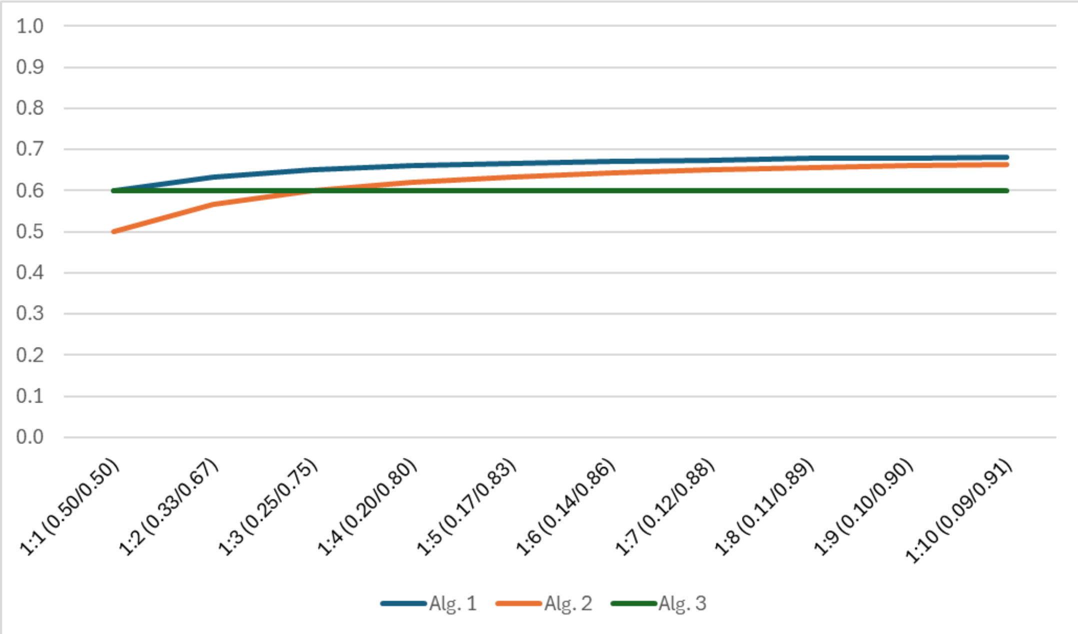


Figure 19: Graphical representation of Table 12.

For the syllable 'ba', Algorithm 1 shows an increase in success score and then stabilizes at 0.700 as the bias becomes more skewed towards high effectiveness. Algorithm 2 demonstrates a similar trend, with its success score increasing slightly and then stabilizing at 0.700. Algorithm 3 maintains a constant success score of 0.600 across all biases, indicating consistent performance regardless of the effectiveness level.

The ANOVA test results for the syllable 'ba' (based on the responses from 10 participants) are as follows:

$$F\text{-}value : 8.31$$

$$P\text{-}value : 0.0015$$

These results indicate a statistically significant difference between the means of the three algorithms. The low p-value suggests that the differences observed are unlikely to be due to chance.

4.2 Comparative and Statistical Analysis

As the aim and a guideline for the result of the analysis of the user study data, the research question should be noted again as the following:

- Is there a significant difference between three different mesh simplification algorithms with two vertex reduction levels (35% and 70%) on the maintenance of the McGurk effect in 3D head models used for lip-sync animation?

In order to answer this question, it can be stated that the comparative analysis of the three syllables reveals distinct trends in the performance of the algorithms. Algorithm 1 exhibits consistent performance for ’va’ but shows a decline for ’ma’ and an increase for ’ba’. Algorithm 2 demonstrates a general decrease in performance for ’va’ but remains stable and slightly improves for ’ma’ and ’ba’. Algorithm 3 consistently performs well for ’va’, remains stable for ’ma’, and shows constant performance for ’ba’. Overall, the results indicate that Algorithm 3 is the most effective in sustaining the McGurk effect across different syllables and biases, followed by Algorithm 1 for ’va’ and ’ba’, and Algorithm 2 for ’ma’ and ’ba’. This suggests that Algorithm 3 provides the best balance of effectiveness and robustness, making it a preferable choice for applications requiring sustained visual illusions.

In conclusion, the comparative analysis and statistical validation highlight Algorithm 3 as the most robust and effective 3D mesh simplification algorithm for sustaining the McGurk effect across different syllables and biases. The fact that such a result was achieved proved the applicability of the concept of "a comparative analysis between mesh simplification algorithms by using a generic methodology", that forms the basis of the research questions. These findings contribute to the broader understanding of the interplay between visual simplification techniques and perceptual phenomena, providing valuable insights for future research and application in visual computing and cognitive science.

CHAPTER 5

DISCUSSION AND CONCLUSION

The purpose of this study was to make a comparative analysis between three specific mesh simplification algorithms; "Quadric Error Metrics" [14], "Triangular Mesh Simplification based on Surface Angle" [15] and "Hierarchical Face Clustering on Polygonal Meshes" [16], by assessing how these algorithms perform against maintaining the McGurk effect at different vertex reduction levels (35% and 70%) of a single head model. This assessment was achieved by analyzing the data gathered from a voluntary user study. In the user study, the participants were exposed to a series of audio-visual stimuli, including high-resolution 3D human head meshes, mouth animations and sounds for , and were tasked with identifying the syllables they perceived. This chapter presents a discussion of the results of this analysis as well as giving conclusive remarks about implications of this study for designers that work on the McGurk effect and for future studies that consider using such utilization of mesh simplification algorithms for 3D models used for McGurk effect.

The first section discusses the methodology used to make the comparative analysis between the three aforementioned mesh simplification algorithms, going in-depth on areas regarding the important aspects in this study. Then it continues with the conclusion section regarding the results of the analysis. Finally, the sections about the implications for both designers and future studies about this topic are offered to conclude this study.

5.1 Discussion

The McGurk effect is a profound phenomenon, even observable across different languages [40]. Thus, it is imperative to assume that the study of this phenomenon can be considered remarkably ubiquitous, especially in recent studies which involve incorporating this auditory-visual illusion in virtual reality applications. Such applications often require intensive use of 3D modeling and a substantial need for high fidelity models, which consume remarkable amounts of storage. In this study, a 3D virtual environment was created to employ the McGurk effect, in order to show that this illusion can be incorporated in virtual reality applications. However, in order to achieve an effective setup, it is important to note that high fidelity 3D models that require high amounts of storage need to be used. This indeed presented a challenge to achieve a balance between a convincing virtual environment for the illusion and an optimal storage space. In regard to this, the importance of the findings of the comparative analysis in this study between the three specific mesh simplification algorithms, in terms of sustaining the McGurk Effect, was substantial.

Overall, the results indicated significant variations in the effectiveness of the McGurk illusion across different syllables regarding the mesh simplification algorithms. In evaluating these simplification algorithms, which were categorized by their effectiveness into medium and high reduction levels, this study provides important insights into the trade-offs between visual detail and auditory accuracy. The algorithms were assessed based on their ability to maintain the McGurk effect while reducing storage consumption, with performance metrics that considered both the fidelity of the visual mesh and the auditory cues' clarity. The result matrices in Chapter 4 for each syllable illustrated how different algorithms performed in maintaining the illusion, with points allocated based on their success in achieving this goal. These findings showed that the role of mesh simplification algorithms becomes more apparent, and correlate with other comparisons, especially in terms of rendering performance [41].

In this study, considering the findings from the comparative analysis, there are areas which invoke further in-depth discussion, such as; seeing different analysis results in each syllable on sustaining the McGurk Effect, how different bias levels affect the analysis results and how this study's analysis projects on other studies.

5.1.1 Different Analysis Results in Syllables on Sustaining the McGurk Effect

Participant data from the user study for each of the three distinct syllables ('va', 'ma', and 'ba') indicated significant variations in the effectiveness of the McGurk illusion across different syllables. For instance, the syllable 'va' showed some susceptibility to the illusion, with only 4 participants (among the total of 42) experiencing the full effect. In contrast, the syllables 'ma' and 'ba' exhibited stronger impacts on perception, with 2 and 10 participants respectively reporting altered auditory perceptions based on visual cues. This variation can be discussed through the lens of phonetics and speech perception, particularly in terms of place of articulation. The place of articulation plays a crucial role in determining the clarity of visual cues during audio-visual integration [26]. Syllables like 'ba' and 'ma' requires the lips coming fully together, creating highly visible movements that are essential for the McGurk effect. In contrast, the syllable 'va' requires contact between the lower lip and the upper teeth, which results in producing subtler visual cues. The dominance of the syllable 'ba' in producing a stronger McGurk effect is in line with the result that Siddig et al. obtained, referring that 'ba' and 'ma' syllables create a stronger McGurk effect comparing to other syllables [5], whereas the reduced visual clarity for 'va' aligns with findings by McGurk et al., who noted that clearer visual articulatory movements enhance the illusion's strength [21]. These findings underscore the complexity and variability of the McGurk effect, highlighting that different phonetic contexts may interact with simplification algorithms in unique ways.

5.1.2 The Effect of Bias

The result matrices in Chapter 4 show the scores for each of the compared algorithms. These scores, and incidentally the resulting graphs, were formed using different bias levels, with the fractions (such as 1:1, 1:2, ..., 1:10) indicating the distribution of points between medium and high effectiveness. Since lesser vertex reduction results in less fidelity loss on the head model, this distribution system is actually a way of seeing the medium and high vertex reduction levels as different difficulties, as in, medium reduction being a medium difficulty and high reduction being a high difficulty, for sustaining the McGurk Effect. As noted in Chapter 3, higher bias levels, such as 1:9 and 1:10, denote giving better scores for high vertex reductions for each particular mesh simplification algorithm. As the bias levels shift from low to high, the resulting graphs showed how different the algorithms compared against each other when medium and high difficulty levels were scored differently.

Observing these resulting graphs, although they showed no remarkable oscillation over varying bias levels, valuing higher vertex reductions with more points resulted in noticeable changes among all syllable categories. This was particularly apparent for the syllables ’va’ and ’ba’, in which there were significant changes of mesh simplification algorithms between low and high bias levels. For example, in the category of the syllable ’ma’, Algorithm 1, which is "Decimate Geometry" implementation of the "Quadric Error Metrics" [14], proved to be better effective at low bias levels but as the high vertex reduction levels were valued more, in terms of sustaining the McGurk Effect, it performed worse than the other two algorithms. Another example such variation is in the category of the syllable ’ba’, where Algorithm 2, which represents "Limited Dissolve" implementation of the "Triangular Mesh Simplification based on Surface Angle" [15], performed seemingly worse at low bias levels but much better at high bias levels.

These findings indicate the importance of making such a comparative analysis using this bias weighting system, as different results were indeed obtained when sustaining the McGurk Effect at higher losses in mesh fidelity were awarded with different points.

5.1.3 Projection of this Study’s Comparison on Other Studies

Three different mesh simplification algorithms were explored in this study: "Quadric Error Metrics" [14], "Triangular Mesh Simplification based on Surface Angle" [15] and "Hierarchical Face Clustering on Polygonal Meshes" [16]. The results of the user study showed that each of these algorithms succeeded in maintaining the McGurk Effect at differing levels. The ANOVA results showed that those differences proved significant; notably, the Hierarchical Face Clustering on Polygonal Meshes algorithm by Garland et al. [16] emerged as the most effective in balancing these factors, offering robust performance even at lower mesh resolutions. This algorithm is known for its efficient edge collapse and hierarchical grouping, which preserved crucial visual features such as mouth shape and articulation while significantly reducing polygon counts.

In this study, mesh simplification algorithms were compared based on their performance at different levels of vertex reduction in sustaining the McGurk effect, which is similar to the method of Garland et al. [14], which focuses on preserving visual quality during simplification. However, this approach differs from other studies such as the one from Cignoni et al. who compared the simplification algorithms using an "empyrical computational cost" between "decimation approaches based on global error evaluation" and "more computationally complex codes based on an energy optimization approach" [8]. This is mainly due to the fact that the comparison in this study was made with a quantitative comparative analysis between different three specific simplification algorithms by assessing their effectiveness on maintaining the McGurk effect on different vertex reduction levels. This unique focus brings together 3D graphics and cognitive science to evaluate how simplifications affect human perception.

Since the main aim is to analyze the maintenance of the McGurk effect on different levels of vertex reductions, this study incorporated a user study to gather data from various participants to make a quantitative comparative analysis between these three algorithms. Specifically, this study sought to assess how different levels of simplified 3D human head meshes, particularly those representing mouth animations aligned with corresponding syllable sounds, impact the reliability and strength of the McGurk illusion.

5.2 Conclusion

The findings of this study affirm the significant impact of mesh simplification on the perception of the McGurk effect. The differential impact observed across syllables highlights that not all simplification algorithms are equally effective in preserving the illusion. The study demonstrates that while some algorithms achieve high levels of effectiveness in sustaining the McGurk illusion, others may fall short, especially when simplifying complex visual details.

Reiterating the research question, the corresponding answer to this question can be summed as the following:

- **Is there a significant difference between three different mesh simplification algorithms with two vertex reduction levels (35% and 70%) on the maintenance of the McGurk effect in 3D head models used for lip-sync animation?**

  In response to this question, it can be stated that the comparative analysis of the three syllables demonstrates distinct trends in the performance of the algorithms. As outlined in Chapter 4, the Hierarchical Face Clustering on Polygonal Meshes algorithm offers the best balance between effectiveness and robustness, making it the preferred choice for applications that require sustained visual illusions. In light of this assessment, based on the sample size and accuracy of the user study and using a high fidelity 3D model to achieve the McGurk effect, it can be shown that there can be a remarkable comparative difference between different mesh simplification algorithms regarding the maintenance of the McGurk effect.

In conclusion, this study underscores the importance of selecting appropriate mesh simplification techniques tailored to the specific requirements of the application. Moreover, this study provides a novel investigation into the effects of 3D mesh simplification on the McGurk effect, contributing to a deeper understanding of how visual fidelity impacts multisensory integration in virtual environments. By demonstrating the performance of different mesh simplification techniques, this study offers valuable insights for designers and developers working with real-time 3D rendering systems, where maintaining the illusion of synchrony between speech and visual cues is critical. The results also highlight the importance of selecting simplification algorithms that provide a balance between visual detail and auditory clarity, with the Hierarchical Face Clustering algorithm standing out as an effective solution for applications that require both performance optimization and perceptual accuracy. The varying effectiveness of the algorithms suggests that a one-size-fits-all approach may not be ideal, and careful consideration must be given to the specific visual and auditory demands of the task at hand.

5.3 Implications for Designers

For designers who work on the McGurk effect in virtual environments, this section provides an insight to the implications regarding this study.

5.3.1 Balancing Realism and Performance

Using this study as a guideline, model designers who work on creating the McGurk effect on virtual environments can benefit in fields such as optimizing visual fidelity on their models while maintaining perceptual accuracy. Because the McGurk effect depends on the synchronization between visual lip movements and auditory speech sounds, in virtual environments, the accuracy of visual representations (such as 3D head meshes) plays a crucial role in the effectiveness of the illusion. However, high-fidelity 3D models can be computationally expensive, especially in real-time applications like VR or AR. Example researches can be shown, in which rendering performance is studied to be increased by using different methods such as "parallel selective rendering" [42].

Mesh simplification algorithms can help reduce the complexity of these 3D models without significantly compromising the visual quality needed for the McGurk effect. This can be especially useful when creating interactive environments where multiple characters or scenes need to be rendered efficiently. By using simplification algorithms that preserve the key features of facial expressions (e.g., the movement of the mouth and lips), designers can maintain perceptual accuracy while optimizing performance.

5.3.2 Improving User Experience in VR/AR

In VR and AR applications, especially those designed to study or simulate the McGurk effect, real-time performance is crucial for a seamless experience, especially considering the role of multisensory feedback and its performance cost [43]. Simplifying meshes can help improve rendering speed and responsiveness, leading to a more fluid and immersive experience for users. This is particularly important when dealing with highly detailed facial animations that can be resource-intensive to render.

Mesh simplification algorithms that preserve critical facial features, such as the mouth, eyes, and eyebrows, can improve frame rates without compromising the visual cues that users rely on for the McGurk effect. This can lead to smoother animations, reducing the risk of lag or visual artifacts that might interfere with the illusion.

5.3.3 Facilitating Large-Scale Simulations

If model designers are working on large-scale simulations or environments with multiple characters, using mesh simplification algorithms can help scale these environments without overburdening the system. For instance, studying the trade-off between system load and the overall quality of a head model and finding a balance between them can lead to how realistic facial avatars can be optimized [44].

In summary, mesh simplification algorithms can greatly benefit model designers working to recreate the McGurk effect in virtual environments by optimizing visual quality, improving real-time performance, and facilitating large-scale simulations. These algorithms help designers balance the need for high visual fidelity with system performance, which is essential for creating effective and immersive environments where the McGurk effect can be explored or leveraged for practical applications.

5.4 Implications for Future Studies

Future research should consider expanding the scope of analysis to include a broader range of syllables and speech sounds to generalize the findings across different phonetic contexts. Additionally, incorporating more diverse participant demographics could provide insights into how individual differences influence the perception of the McGurk effect with simplified 3D meshes.

Further studies, especially for developers, could also explore the integration of advanced algorithms that combine multiple factors, such as real-time adaptability to varying levels of mesh detail or the incorporation of machine learning techniques to optimize simplification processes. Additionally, for designers and animators, examining the impact of different types of animations and textures on the McGurk effect could offer a more comprehensive understanding of how visual and auditory cues interact in 3D environments.

It should also be noted that the McGurk effect is mostly achieved through the perception of the mouth and nasal areas of a human face. And since this study is made using the mesh simplification algorithms being utilized on the whole head mesh, some of the optimization, which occurs, for example, on the back of the head or on the forehead or the neck areas, might not have any effect on the illusion but still considered in the vertex count reduction. A future study could focus solely on the most prominent areas of the face in terms of utilizing the mesh simplification algorithms.

There is also a research about how music-trained people impacted differently by the McGurk effect, published in 2024 [45]. This could also indicate that the demographic scope of this study can be narrowed down, also by checking each participant's experience in music.

While the most popular mesh simplification algorithms represent significant strides in 3D mesh optimization, challenges remain in balancing simplification with preserving semantic and perceptual fidelity. Future research directions may explore adaptive strategies that dynamically adjust mesh complexity based on user-defined criteria or contextual requirements, advancing the capabilities of interactive graphics and virtual environments.

Ultimately, this study paves the way for more refined approaches to mesh simplification in applications where the accurate perception of visual and auditory stimuli is essential, contributing valuable insights to both theoretical and practical aspects of 3D modeling and human-computer interaction.

REFERENCES

[1] N. Joncic, *3D Digital Recording of Archaeological, Architectural and Artistic Heritage*. University of Ljubljana, 03 2020.

[2] M. Botsch, M. Pauly, C. Rössl, S. Bischoff, and L. Kobbelt, "Geometric modeling based on triangle meshes," in *ACM SIGGRAPH 2006 Courses*, pp. 1–es, 2006.

[3] G. M. Clark, "Personal reflections on the multichannel cochlear implant and a view of the future," *Journal of Rehabilitation Research & Development*, vol. 45, no. 5, 2008.

[4] R. A. Potamias, S. Ploumpis, and S. Zafeiriou, "Neural mesh simplification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18583–18592, 2022.

[5] A. Siddig, P. W. Sun, M. Parker, and A. Hines, "Perception deception: Audio-visual mismatch in virtual reality using the mcgurk effect," in *AICS 2019*, pp. 176–187, 2019.

[6] V. Kasavajhala, "Solid state drive vs. hard disk drive price and performance study," 2011.

[7] K. L. Low and T. S. Tan, "Model simplification using vertex-clustering," in *Proceedings of the 1997 symposium on Interactive 3D graphics*, pp. 75–ff, April 1997.

[8] P. Cignoni, C. Montani, and R. Scopigno, "A comparison of mesh simplification algorithms," *Computers & Graphics*, vol. 22, no. 1, pp. 37–54, 1998.

[9] X. Luan, Y. Xie, L. Ying, and L. Wu, "Research and development of 3d modeling," *IJCSNS International Journal of Computer Science and Network Security*, vol. 8, no. 1, pp. 49–53, 2008.

[10] R. Ohbuchi, H. Masuda, and M. Aono, "Embedding data in 3d models," in *Interactive Distributed Multimedia Systems and Telecommunication Services: 4th International Workshop, IDMS'97 Darmstadt, Germany, September 10–12, 1997 Proceedings 4*, pp. 1–10, Springer Berlin Heidelberg, 1997.

[11] D. Ostrovka, T. Teslyuk, P. Veselý, and I. Prots'ko, "The analysis and comparison of file formats for the construction of ios os three-dimensional objects for augmented reality systems," in *DCSmart*, pp. 325–334, 2019.

[12] Y. Wang, J. Zheng, and H. Wang, "Fast mesh simplification method for three-dimensional geometric models with feature-preserving efficiency," *Scientific Programming*, vol. 2019, no. 1, p. 4926190, 2019.

[13] J. H. Clark, "Hierarchical geometric models for visible surface algorithms," *Communications of the ACM*, vol. 19, no. 10, pp. 547–554, 1976.

[14] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 209–216, ACM, 1997.

[15] H. L. Li and S. Kim, "Triangular mesh simplification based on surface angle," in *MITA2019*, pp. 143–146, 2019.

[16] M. Garland, A. Willmott, and P. S. Heckbert, "Hierarchical face clustering on polygonal surfaces," in *Proceedings of the 2001 symposium on Interactive 3D graphics*, pp. 49–58, ACM, 2001.

[17] J. Z. Li, *CAD, 3D Modeling, engineering analysis, and prototype experimentation*. Imprint: Springer: Springer International Publishing, 2015.

[18] J. Keil, D. Edler, T. Schmitt, and F. Dickmann, "Creating immersive virtual environments based on open geospatial data and game engines," *KN-Journal of Cartography and Geographic Information*, vol. 71, no. 1, pp. 53–65, 2021.

[19] Y. Hendriyani and V. A. Amrizal, "The comparison between 3D Studio Max and Blender based on software qualities," *Journal of Physics: Conference Series*, vol. 1387, p. 012030, November 2019.

[20] D. Todorović, "What are visual illusions?," *Perception*, vol. 49, no. 11, pp. 1128–1199, 2020.

[21] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, 1976.

[22] B. Allahgholipour, "Multi-sensory perception in virtual world animation using mcgurk effect," Master's thesis, Middle East Technical University, 2019.

[23] A. Wlaszczyk, "Time and uncertainty in computational models of audio-visual speech integration," 2024.

[24] D. Burnham and B. Dodd, "Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the mcgurk effect," *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, vol. 45, no. 4, pp. 204–220, 2004.

[25] R. Thézé, M. A. Gadiri, L. Albert, A. Provost, A.-L. Giraud, and P. Mégevand, "Animated virtual characters to explore audio-visual speech in controlled and naturalistic environments," *Scientific reports*, vol. 10, no. 1, p. 15540, 2020.

[26] K. J. Van Engen, A. Dey, M. S. Sommers, and J. E. Peelle, "Audiovisual speech perception: Moving beyond mcgurk," *The journal of the acoustical society of america*, vol. 152, no. 6, pp. 3216–3225, 2022.

[27] S. Walker, V. Bruce, and C. O'Malley, "Facial identity and facial speech processing: Familiar faces and voices in the mcgurk effect," *Perception & psychophysics*, vol. 57, no. 8, pp. 1124–1133, 1995.

[28] B. Fang, X. Li, G. Han, and J. He, "Facial expression recognition in educational research from the perspective of machine learning: A systematic review," *IEEE Access*, 2023.

[29] L. Bajrami and M. Ismaili, "The role of video materials in efl classrooms," *Procedia-Social and Behavioral Sciences*, vol. 232, pp. 502–506, 2016.

[30] Y. Hayashi and K. Sekiyama, "Native-foreign language effect in the mcgurk effect: a test with chinese and japanese," *perception*, vol. 1, p. 2, 1998.

[31] R. J. Hirst, J. E. Stacey, L. Cragg, P. C. Stacey, and H. A. Allen, "The threshold for the mcgurk effect in audio-visual noise decreases with development," *Scientific Reports*, vol. 8, no. 1, p. 12372, 2018.

[32] K. Chládková, V. J. Podlipskỳ, N. Nudga, and Š. Šimáčková, "The mcgurk effect in the time of pandemic: age-dependent adaptation to an environmental loss of visual speech cues," *Psychonomic Bulletin & Review*, vol. 28, pp. 992–1002, 2021.

[33] P. Ladefoged, K. Johnson, and P. Ladefoged, *A Course in Phonetics*. Thomson Wadsworth, 3 ed., 2006.

[34] S. Sumpeno, M. Hariadi, and A. M. Syarif, "Development of indonesian text-to-audiovisual synthesis system using syllable concatenation approach to support indonesian learning," *International Journal of Emerging Technologies in Learning*, vol. 12, no. 2, 2017.

[35] B. Foundation, "Bmesh decimator that uses an edge collapse method." `https://github.com/blender/blender/blob/main/source/blender/bmesh/tools/bmesh_decimate_collapse.cc`, 2024. Accessed: 2024-07-15.

[36] G. Kalton and I. Flores-Cervantes, "Weighting methods," *Journal of Official Statistics*, vol. 19, no. 2, p. 81, 2003.

[37] R. A. Fisher, "Statistical methods for research workers," in *Breakthroughs in Statistics: Methodology and Distribution*, pp. 66–70, New York, NY: Springer New York, 1970.

[38] J. F. Magnotti, A. Lado, Y. Zhang, A. Maasø, A. Nath, and M. S. Beauchamp, "Repeatedly experiencing the mcgurk effect induces long-lasting changes in auditory speech perception," *Communications Psychology*, vol. 2, no. 1, p. 25, 2024.

[39] J. Stacey, C. Howard, S. Mitra, and P. Stacey, "The relationship between eye movements and the mcgurk effect when stimuli are presented in noise," 2017.

[40] A. D. Solarte, "The mcgurk effect across languages," *Eureka*, vol. 8, no. 1, 2023.

[41] H. Li and S. Kim, "Mesh simplification algorithms for rendering performance," *International Journal of Engineering Research and Technology*, vol. 13, no. 6, pp. 1110–1119, 2020.

[42] K. Debattista, A. Chalmers, R. Gillibrand, P. Longhurst, G. Mastoropoulou, and V. Sundstedt, "Parallel selective rendering of high-fidelity virtual environments," *Parallel Computing*, vol. 33, no. 6, pp. 361–376, 2007.

[43] K. Apostolou and F. Liarokapis, "A systematic review: The role of multisensory feedback in virtual reality," in *2022 IEEE 2nd International Conference on Intelligent Reality (ICIR)*, pp. 39–42, 2022.

[44] B. Liu, "Optimizing realistic 3d facial models for vr avatars through mesh simplification," 2023.

[45] H.-H. Lee, K. Groves, P. Ripollés, and M. Carrasco, "Audiovisual integration in the mcgurk effect is impervious to music training," *Scientific Reports*, vol. 14, no. 1, p. 3262, 2024.

APPENDIX A

ETHICAL CLEARANCE

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

05 TEMMUZ 2024

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Doç. Dr. Saniye Tuğba TOKEL

Danışmanlığını yürüttüğünüz Ümit ERONAT'ın ***"Sanal dünya animasyonunda yanılsama algısının McGurk etkisi ve 3 boyutlu nesne örgüsü optimizasyonuna dayalı analizi"*** başlıklı araştırmanız İnsan Araştırmaları Etik Kurulu tarafından uygun görülerek **0418-ODTUİAEK-2024** protokol numarası ile onaylanmıştır

Bilgilerinize saygılarımla sunarım

Prof. Dr. Ş. Halil TURAN
Başkan

Prof.Dr. İ. Semih AKÇOMAK
Üye

Doç. Dr. Ali Emre Turgut
Üye

Doç. Dr. Şerife SEVİNÇ
Üye

Doç.Dr. Murat Perit ÇAKIR
Üye

Dr. Öğretim Üyesi Süreyya ÖZCAN KABASAKAL
Üye

Dr. Öğretim Üyesi Müge GÜNDÜZ
Üye

APPENDIX B

CONSENT FORM

Çalışmanın Amacı Nedir?
Araştırmanın amacı, bir görsel ilüzyonun sanal ortamda etki düzeyinin incelenmesidir.

Bize Nasıl Yardımcı Olmanızı İsteyeceğiz?
Araştırmaya katılmayı kabul ederseniz, size bir dizi çoktan seçmeli soru sorulacaktır. Her soruda birkaç saniyelik kısa videolar izlemeniz ve videolarda hangi sesi duyduğunuzu videoların altında bulunan seçeneklerden birini işaretleyerek değerlendirmeniz beklenmektedir. Bu ankete katılım yaklaşık 8-10 dakika sürecektir.

Sizden Topladığımız Bilgileri Nasıl Kullanacağız?
Bu çalışmaya katılmak tamamen gönüllülük esasına dayalıdır. Herhangi bir yaptırıma veya cezaya maruz kalmadan çalışmaya katılmayı reddedebilir veya çalışmayı bırakabilirsiniz. Ancak araştırmanın tamamlanabilmesi için tüm sorulara cevap vermeniz beklenmektedir, aksi takdirde katılımınız değerlendirilmeyecektir. Ankette sizden kimlik veya kurum belirleyici hiçbir bilgi istenmeyecektir. Eposta bilgileriniz herhangi bir şekilde kaydedilmeyecek, ankete katılımınız anonim kalacaktır. Cevaplarınız tamamıyla gizli tutulacak ve sadece araştırmacılar tarafından değerlendirilecektir. Katılımcılardan elde edilecek bilgiler toplu halde değerlendirilecek ve bilimsel yayımlarda kullanılacaktır.

Katılımınızla İlgili Bilmeniz Gerekenler
Çalışma, genel olarak kişisel rahatsızlık verecek sorular içermemektedir. Her bir anket sorusunda kısa videolar izletileceğinden ve duyduğunuz sesi değerlendirmeniz istendiğinden dolayı, görsel ve işitsel bir rahatsızlığınız varsa, katılım sırasında sizi rahatsız edecek herhangi bir durumla karşılaşırsanız dilediğiniz gibi cevaplama işini yarıda bırakıp çıkmakta serbestsiniz. Böyle bir durumda çalışmanız değerlendirilmeyecektir. En iyi çalışma kalitesi için kulaklık veya ses kalitesi iyi olan hoparlörler kullanmanız önerilir.

Araştırmayla İlgili Daha Fazla Bilgi Almak İsterseniz...
Bu çalışmaya katıldığınız için şimdiden teşekkür ederiz. Çalışma hakkında daha fazla bilgi almak için iletişim kurabilirsiniz.

Yukarıdaki bilgileri okudum ve bu çalışmaya tamamen gönüllü olarak katılıyorum.

- Kabul ediyorum
- Kabul etmiyorum

APPENDIX C

QUESTIONNAIRE

Kişisel bilgiler

1. Yaş aralığınız nedir?

- 18-35
- 36-50
- 51-70
- 71+

2. Görmeyi veya işitmeyi üst düzeyde etkileyen bir hastalığınız, bir aksaklık durumunuz var mı? (Ör. Görüşünüz bir hastalık sebebiyle %100 olmayabilir, bir kulak-burun-boğaz hastalığı sonucu işitme cihazı kullanıyor olabilirsiniz, vs.)

- Hayır, yok
- Görmeyi üst düzeyde etkileyen bir durumum var
- İşitmeyi üst düzeyde etkileyen bir durumum var
- Hem görme hem de işitme konusunda aksaklığa sebep olan durumlarım var

İşitme kontrolu 1

Bu sorunun cevabı ankette vereceğiniz diğer cevapların analizinde kullanılacaktır. Lütfen dikkatlice dinleyip hangi sesi duyuyorsanız cevaplamaya çalışın.

Hangi sesi duyuyorsunuz?

- pa
- ma
- va
- ba
- hiçbiri

İşitme kontrolu 2

Bu sorunun cevabı ankette vereceğiniz diğer cevapların analizinde kullanılacaktır. Lütfen dikkatlice dinleyip hangi sesi duyuyorsanız cevaplamaya çalışın.

Hangi sesi duyuyorsunuz?

- ma
- pa
- ba
- va
- hiçbiri

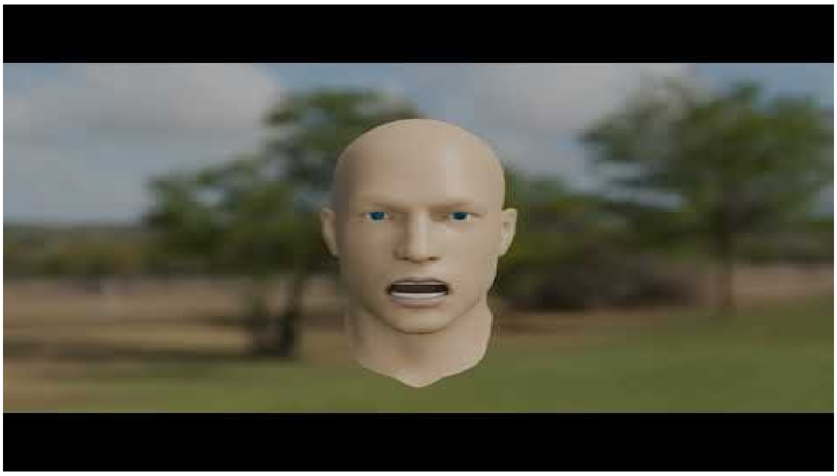
İşitme kontrolu 3

Bu sorunun cevabı ankette vereceğiniz diğer cevapların analizinde kullanılacaktır. Lütfen dikkatlice dinleyip hangi sesi duyuyorsanız cevaplamaya çalışın.

Hangi sesi duyuyorsunuz?

- ba
- va
- ga
- fa
- hiçbiri

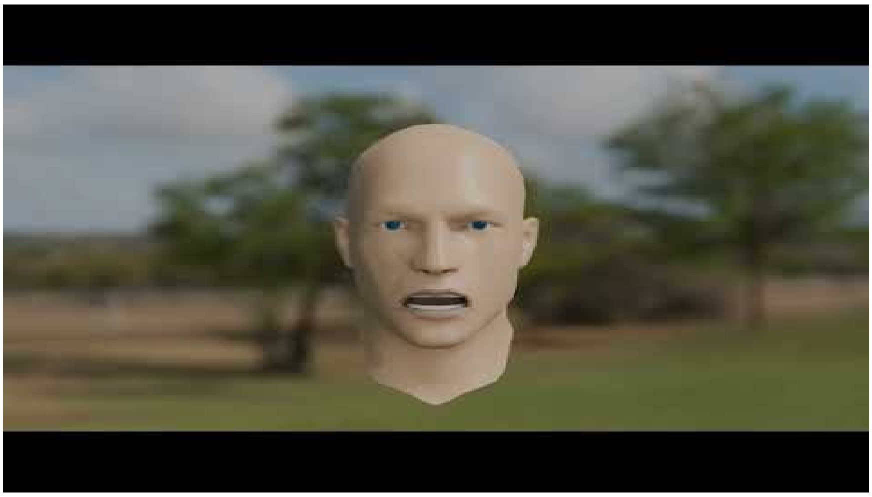
Soru 1



Hangi sesi duyuyorsunuz?

- va
- ba
- ma
- fa
- hiçbiri

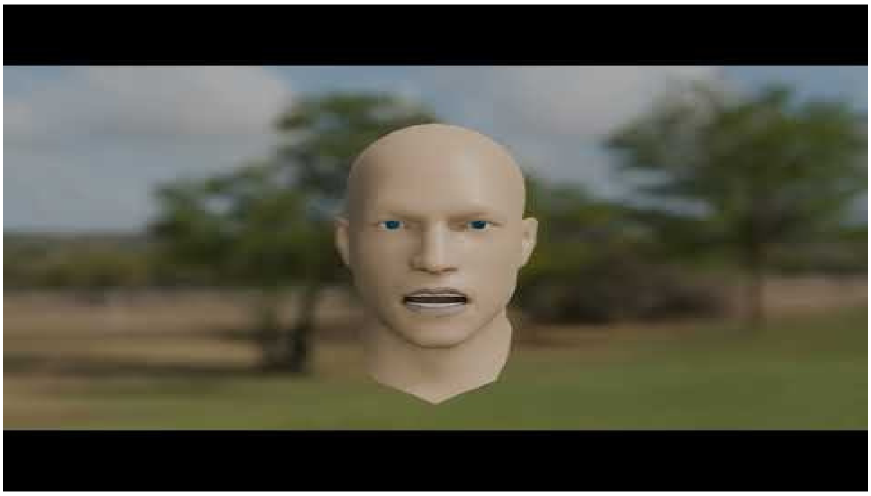
Soru 2



Hangi sesi duyuyorsunuz?

- va
- ma
- fa
- ba
- hiçbiri

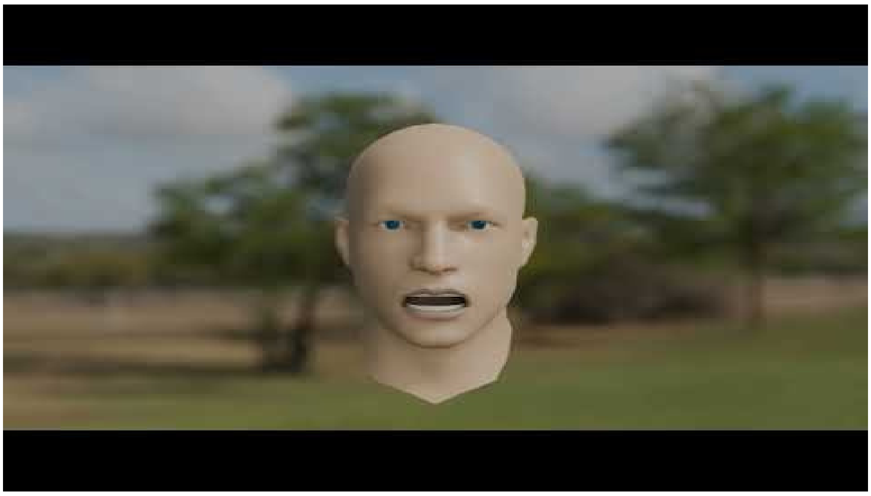
Soru 3



Hangi sesi duyuyorsunuz?

- fa
- ma
- ba
- va
- hiçbiri

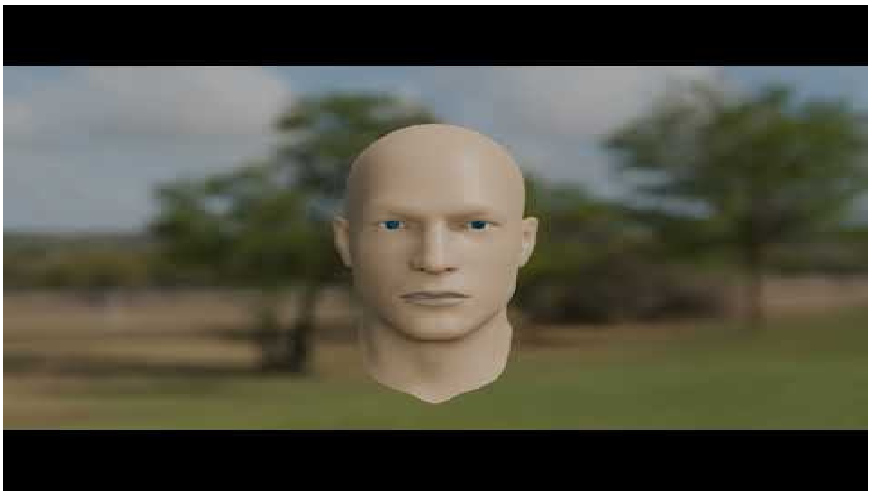
Soru 4



Hangi sesi duyuyorsunuz?

- ma
- fa
- va
- ba
- hiçbiri

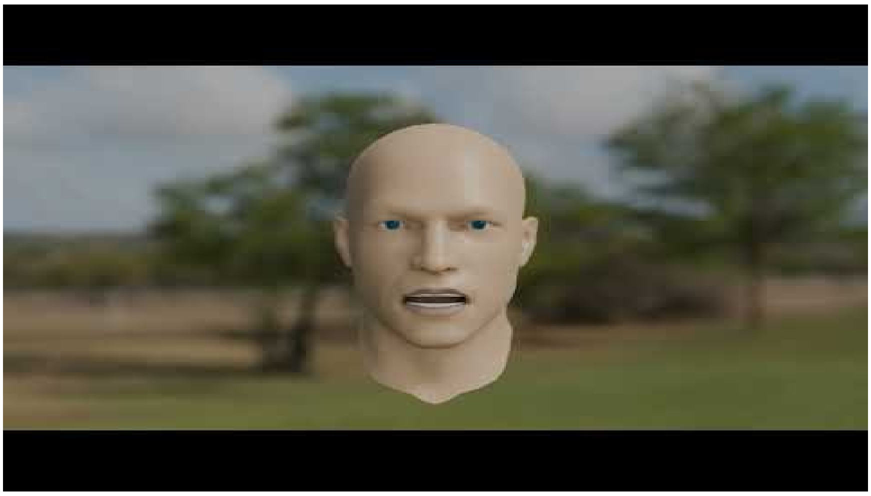
Soru 5



Hangi sesi duyuyorsunuz?

- ma
- fa
- ba
- va
- hiçbiri

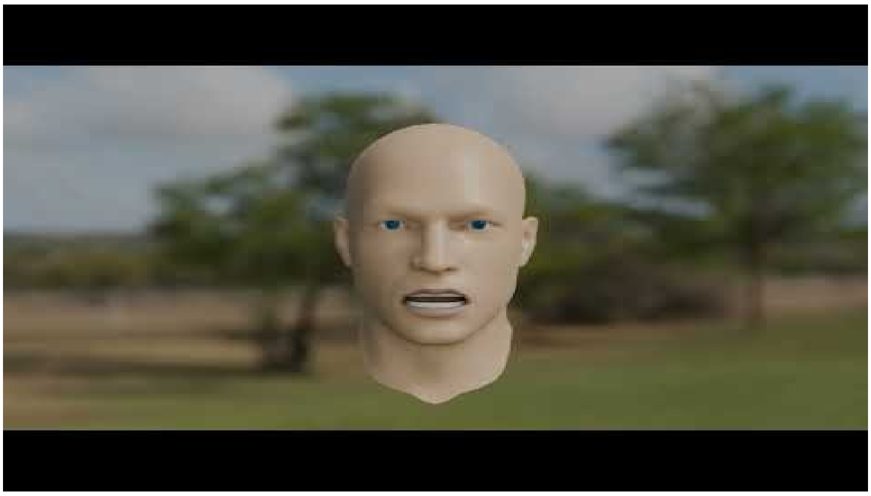
Soru 6



Hangi sesi duyuyorsunuz?

- va
- ga
- pa
- ma
- hiçbiri

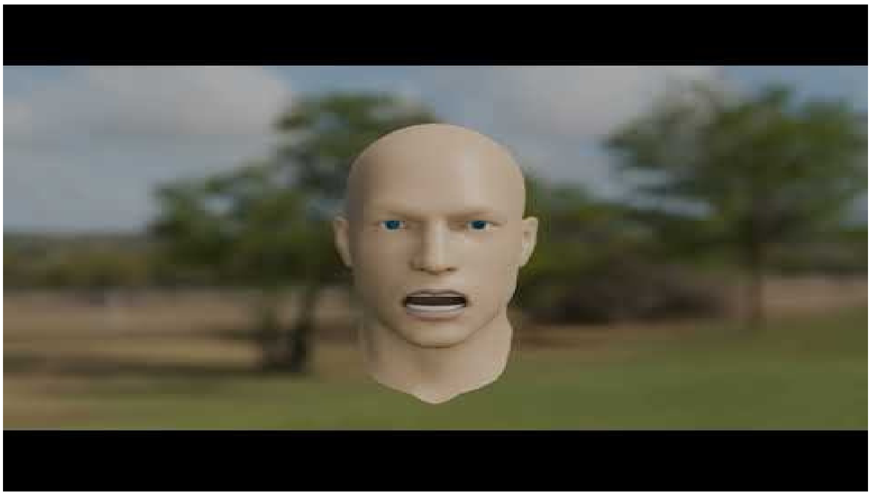
Soru 7



Hangi sesi duyuyorsunuz?

- ma
- ba
- fa
- va
- hiçbiri

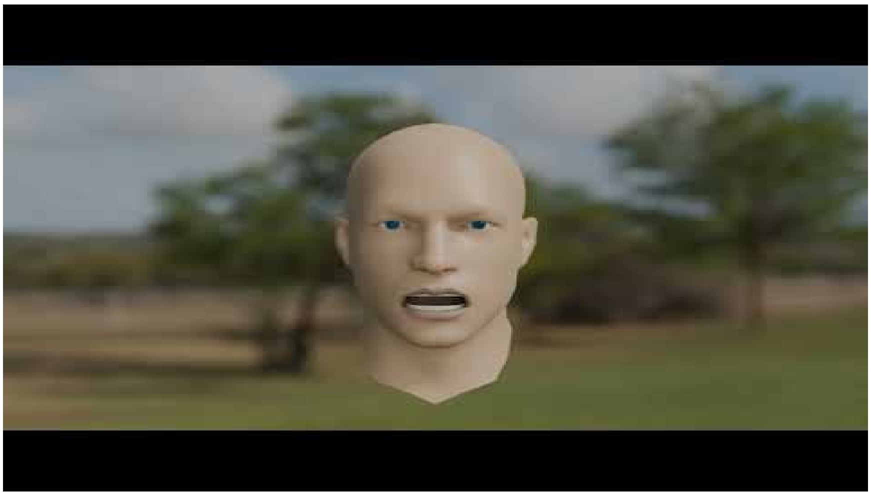
Soru 8



Hangi sesi duyuyorsunuz?

- ba
- fa
- va
- ma
- hiçbiri

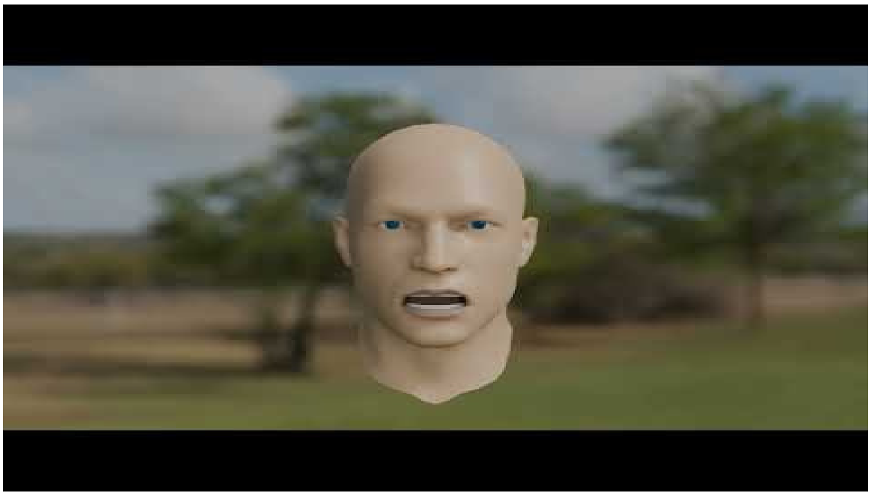
Soru 9



Hangi sesi duyuyorsunuz?

- ma
- fa
- va
- ba
- hiçbiri

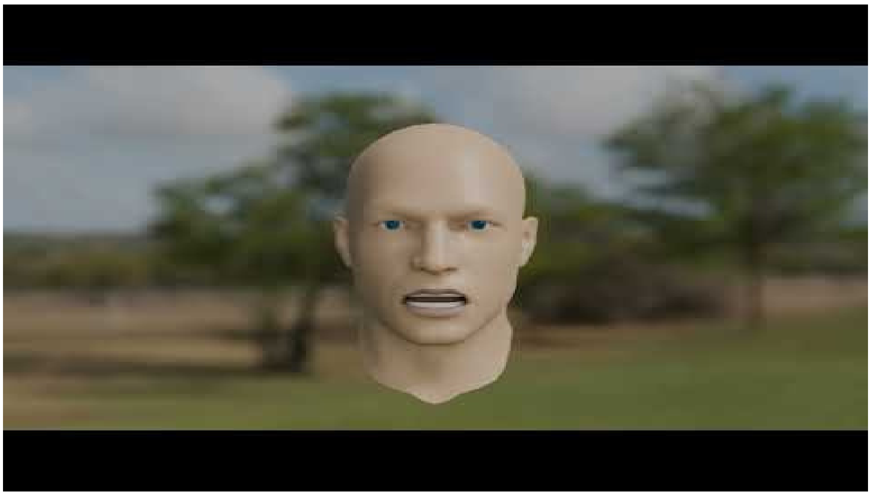
Soru 10



Hangi sesi duyuyorsunuz?

- ma
- va
- ba
- fa
- hiçbiri

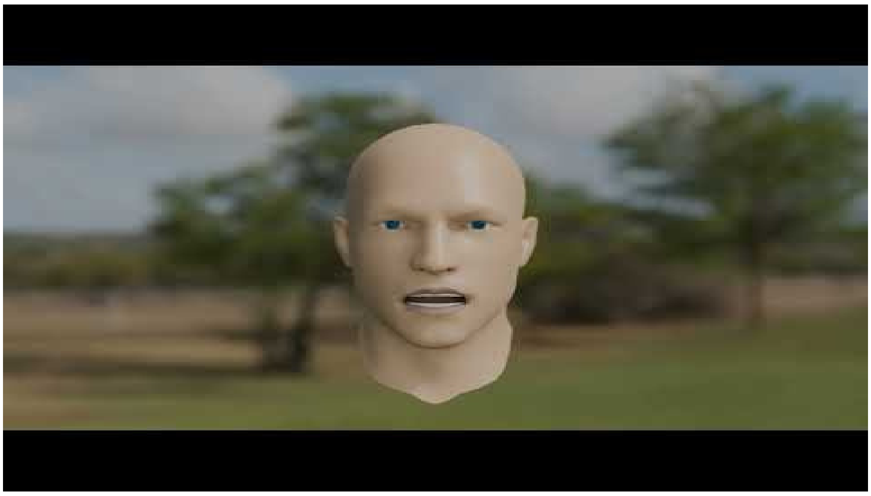
Soru 11



Hangi sesi duyuyorsunuz?

- ba
- va
- ma
- fa
- hiçbiri

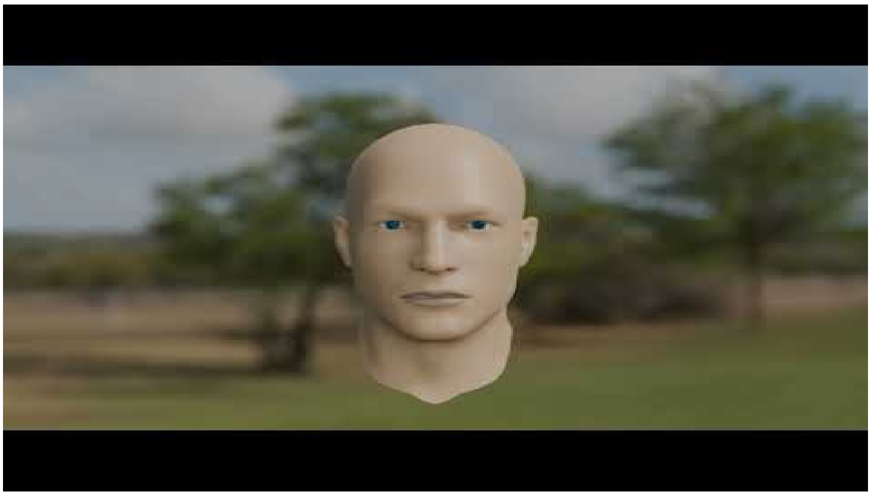
Soru 12



Hangi sesi duyuyorsunuz?

- ma
- ba
- va
- fa
- hiçbiri

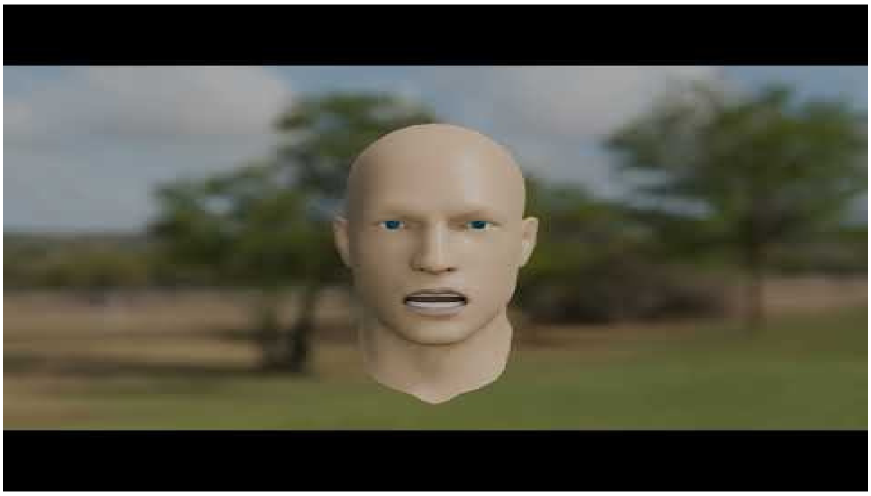
Soru 13



Hangi sesi duyuyorsunuz?

- fa
- va
- ma
- ba
- hiçbiri

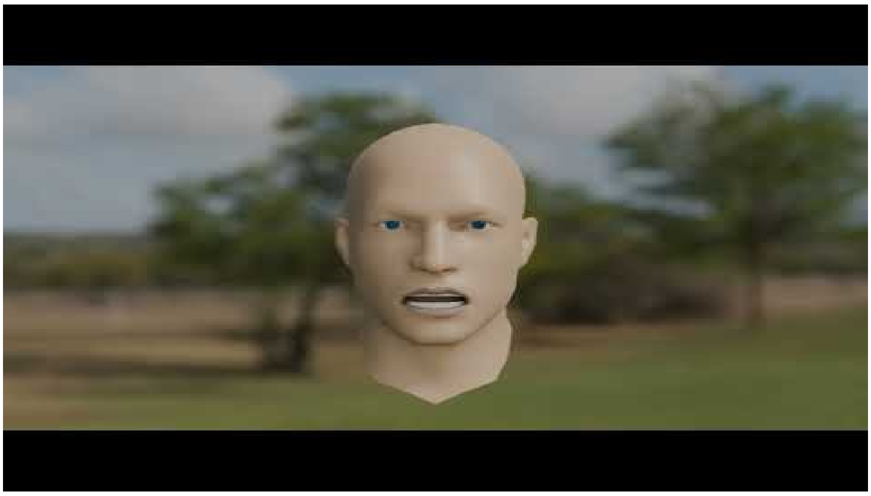
Soru 14



Hangi sesi duyuyorsunuz?

- pa
- ba
- fa
- va
- hiçbiri

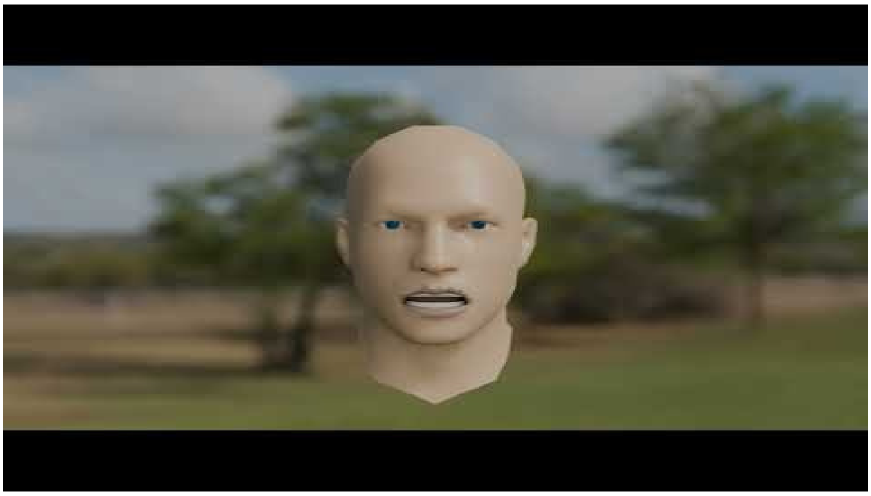
Soru 15



Hangi sesi duyuyorsunuz?

- pa
- ba
- fa
- va
- hiçbiri

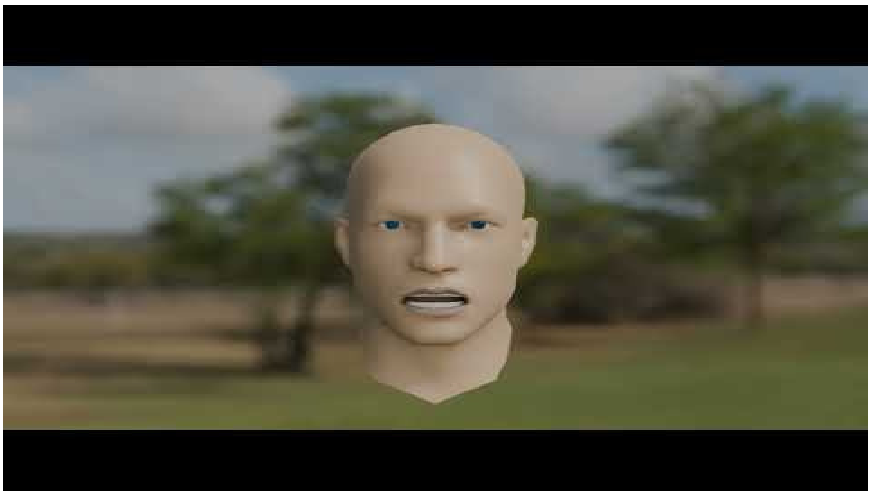
Soru 16



Hangi sesi duyuyorsunuz?

- ba
- pa
- va
- fa
- hiçbiri

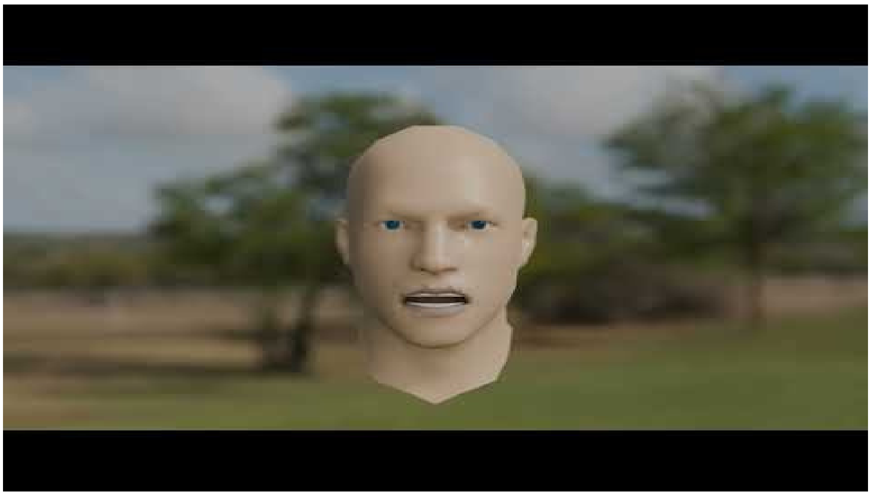
Soru 17



Hangi sesi duyuyorsunuz?

- va
- ma
- fa
- ba
- hiçbiri

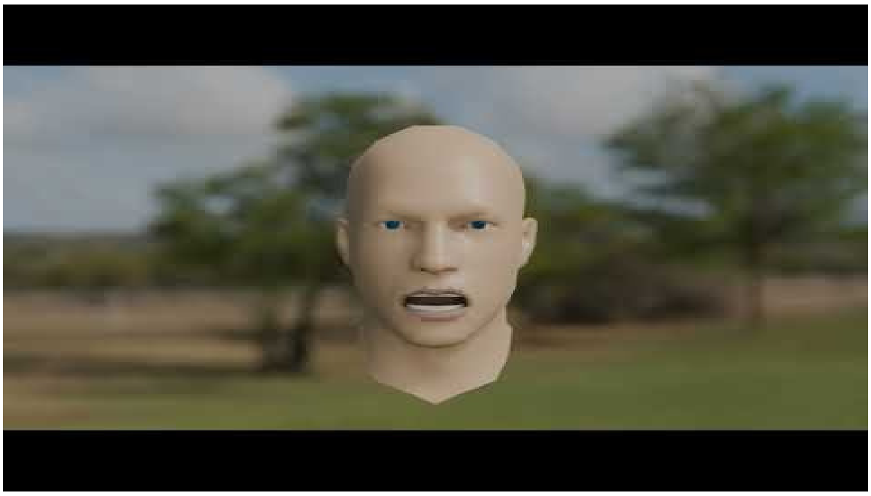
Soru 18



Hangi sesi duyuyorsunuz?

- va
- ma
- ba
- fa
- hiçbiri

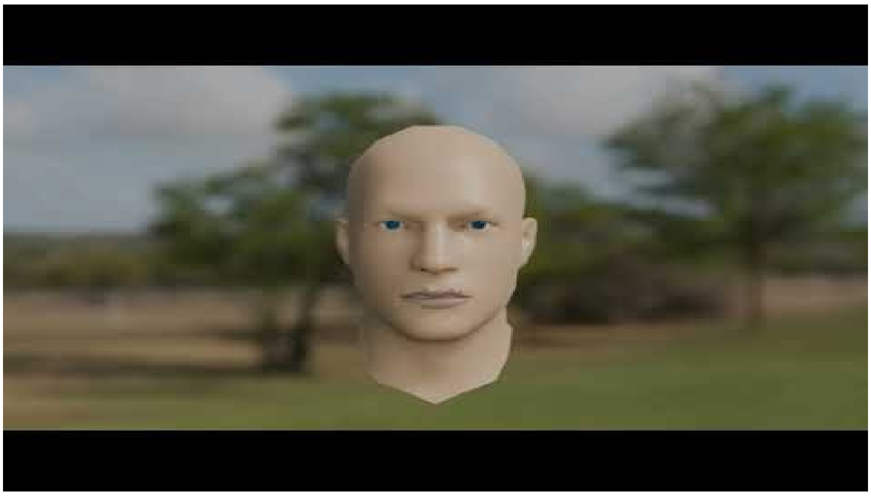
Soru 19



Hangi sesi duyuyorsunuz?

- pa
- va
- fa
- ba
- hiçbiri

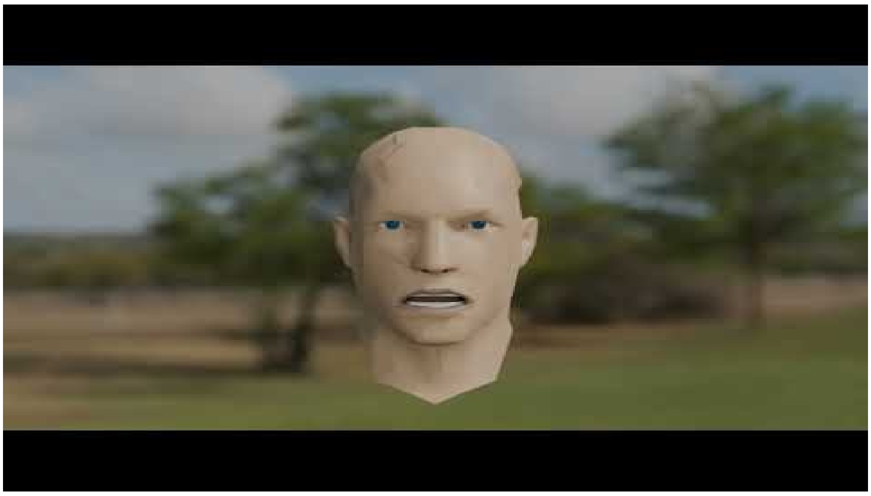
Soru 20



Hangi sesi duyuyorsunuz?

- ba
- ma
- pa
- va
- hiçbiri

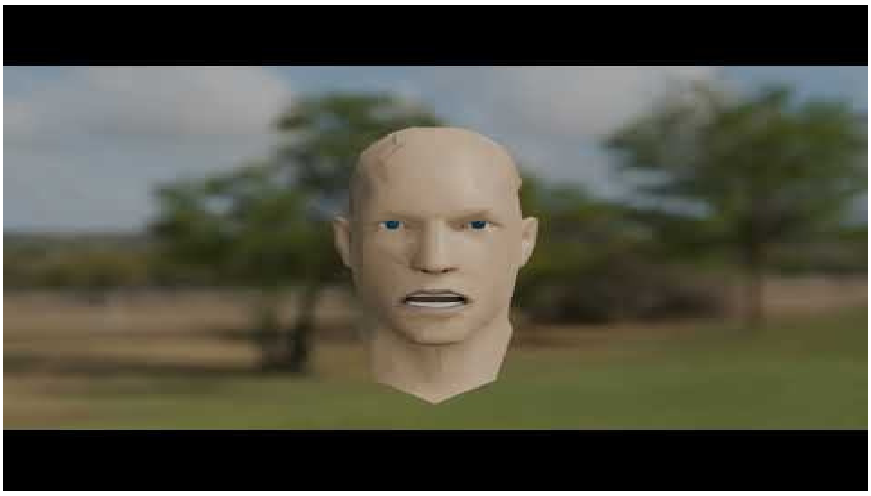
Soru 21



Hangi sesi duyuyorsunuz?

- ba
- ma
- va
- fa
- hiçbiri

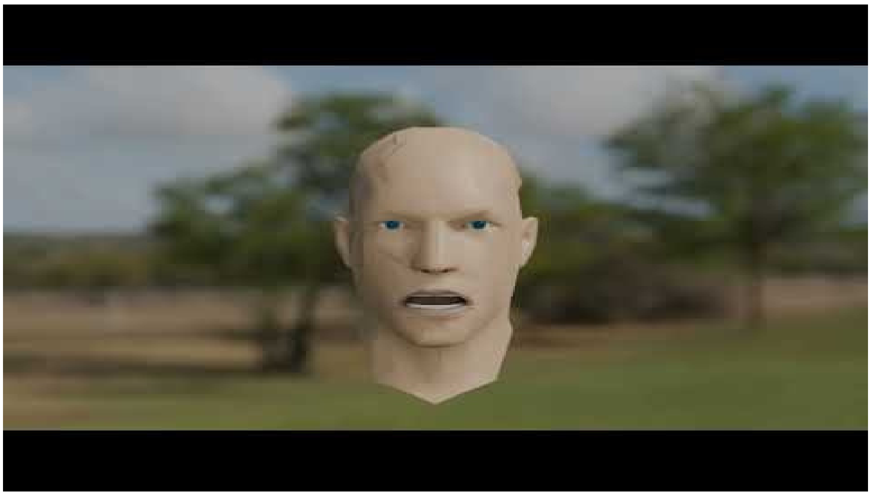
Soru 22



Hangi sesi duyuyorsunuz?

- pa
- ba
- ma
- va
- hiçbiri

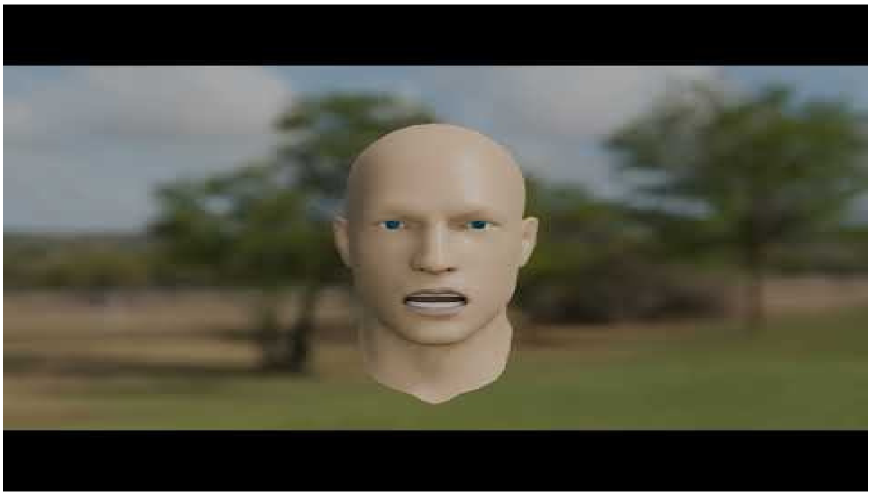
Soru 23



Hangi sesi duyuyorsunuz?

- pa
- ma
- va
- fa
- hiçbiri

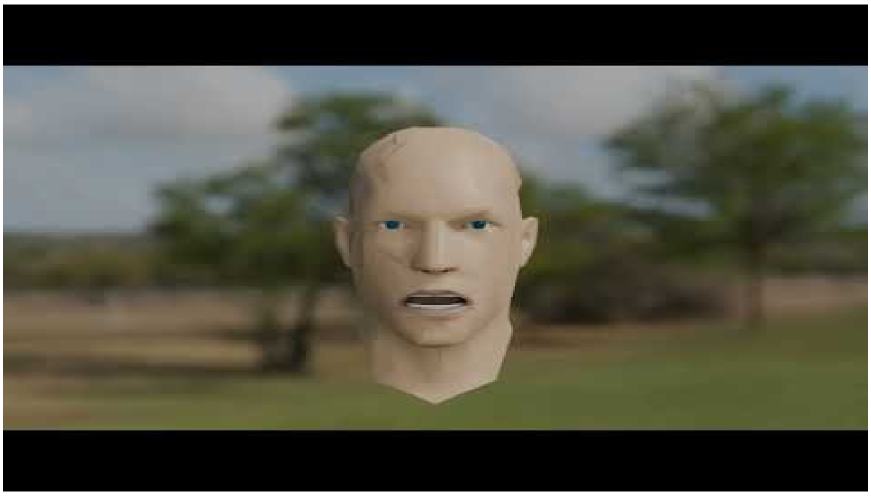
Soru 24



Hangi sesi duyuyorsunuz?

- ma
- ba
- va
- pa
- hiçbiri

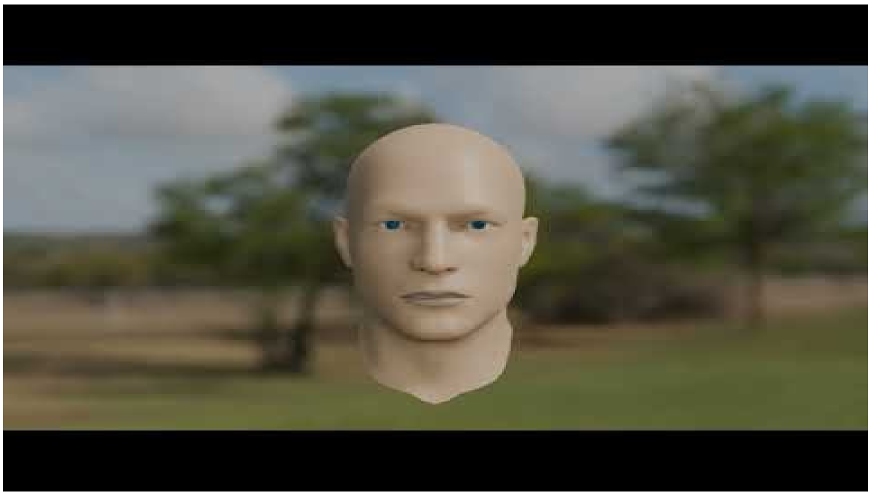
Soru 25



Hangi sesi duyuyorsunuz?

- ba
- va
- fa
- ma
- hiçbiri

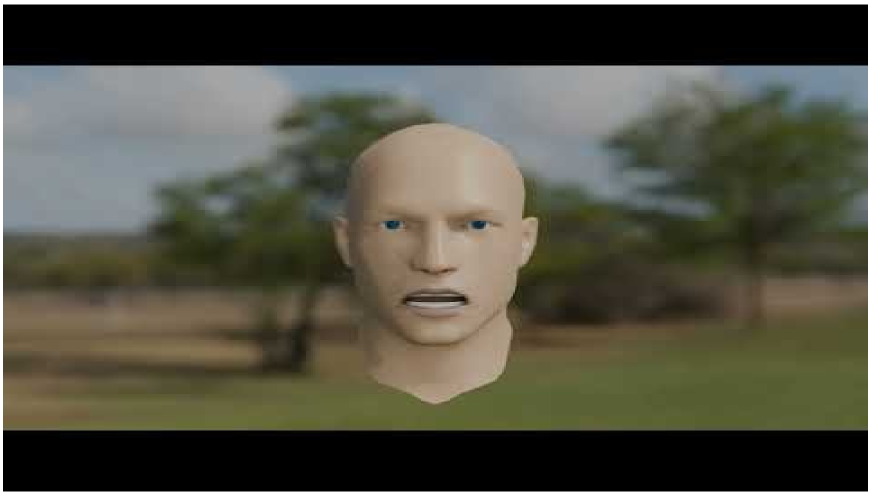
Soru 26



Hangi sesi duyuyorsunuz?

- pa
- ba
- va
- fa
- hiçbiri

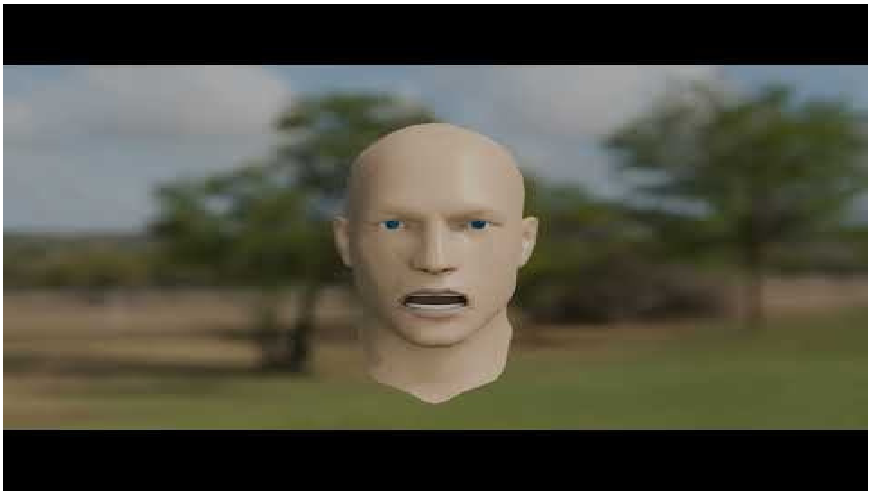
Soru 27



Hangi sesi duyuyorsunuz?

- ba
- pa
- va
- fa
- hiçbiri

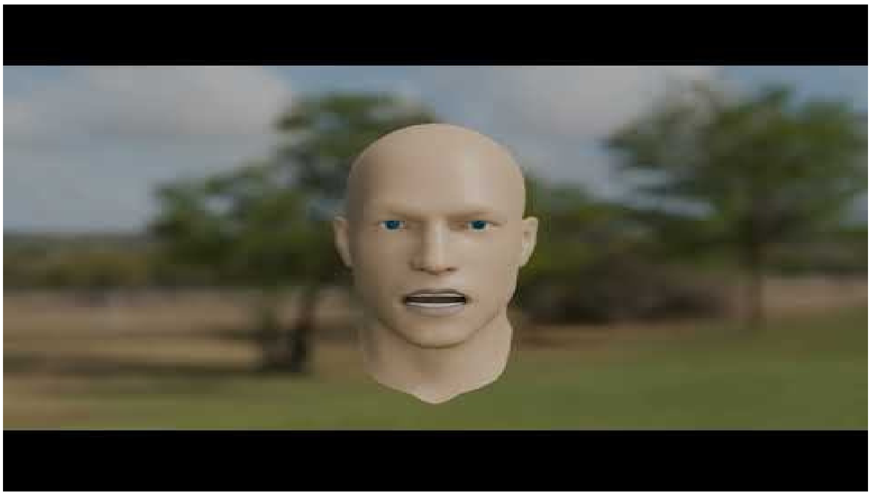
Soru 28



Hangi sesi duyuyorsunuz?

- va
- fa
- ma
- ba
- hiçbiri

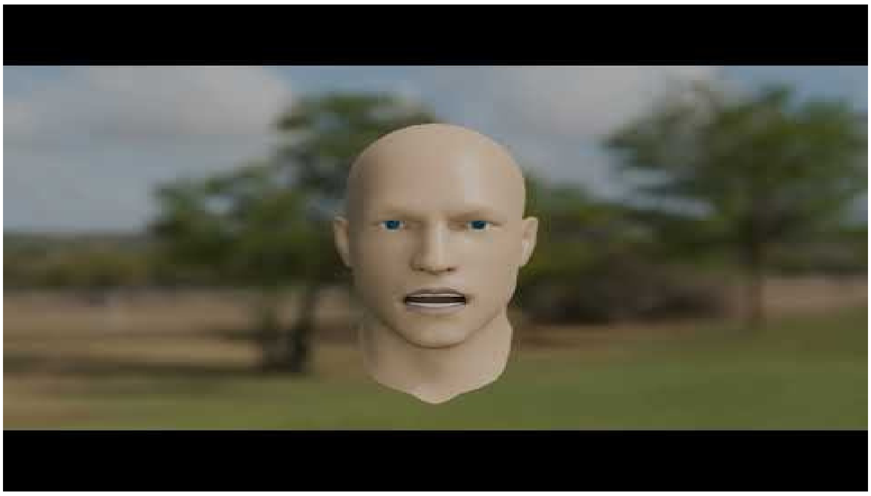
Soru 29



Hangi sesi duyuyorsunuz?

- pa
- ba
- fa
- va
- hiçbiri

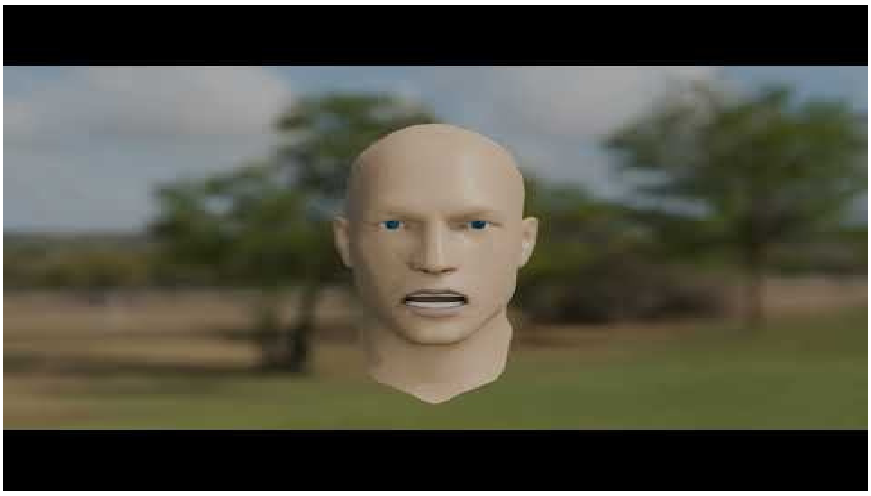
Soru 30



Hangi sesi duyuyorsunuz?

- va
- fa
- ba
- pa
- hiçbiri

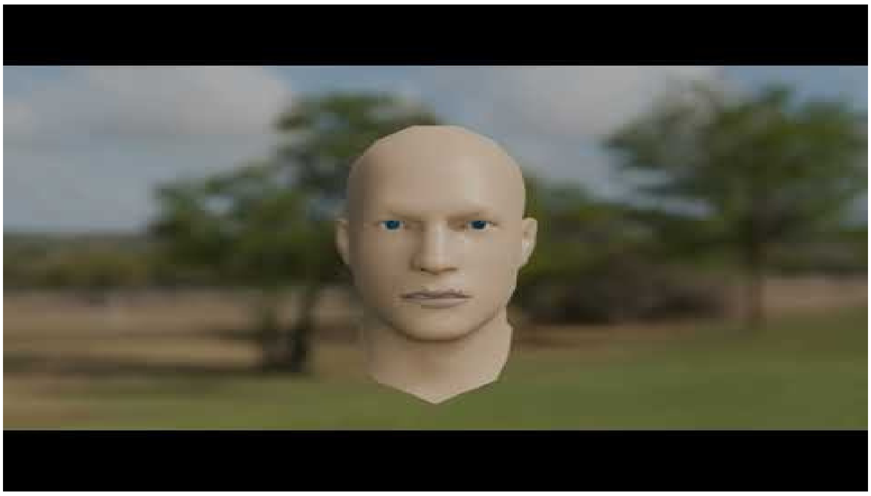
Soru 31



Hangi sesi duyuyorsunuz?

- va
- ma
- ba
- fa
- hiçbiri

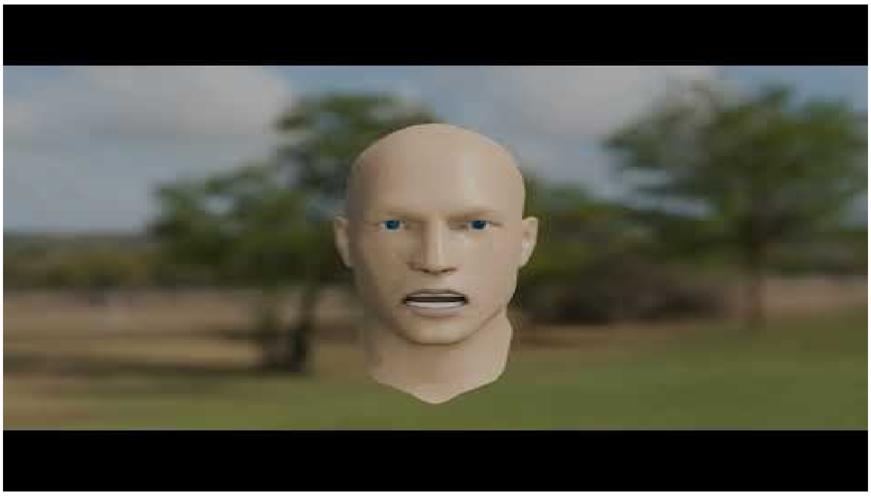
Soru 32



Hangi sesi duyuyorsunuz?

- fa
- ba
- va
- ma
- hiçbiri

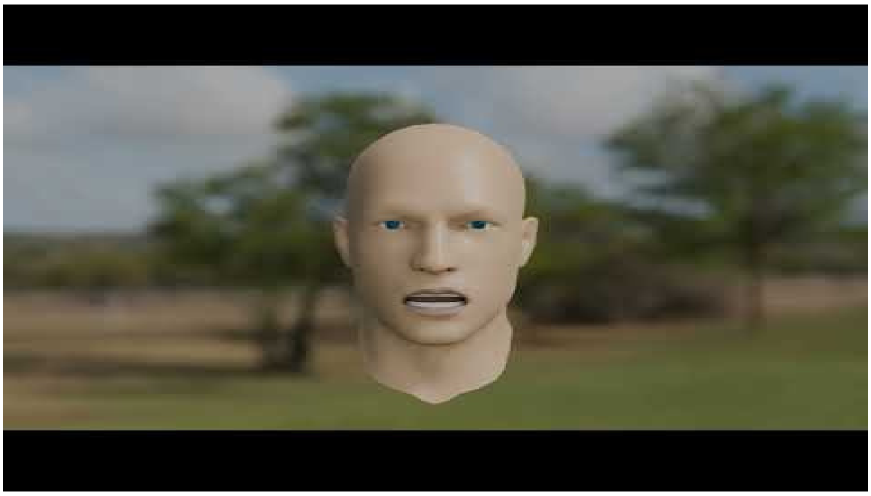
Soru 33



Hangi sesi duyuyorsunuz?

- va
- ba
- ma
- fa
- hiçbiri

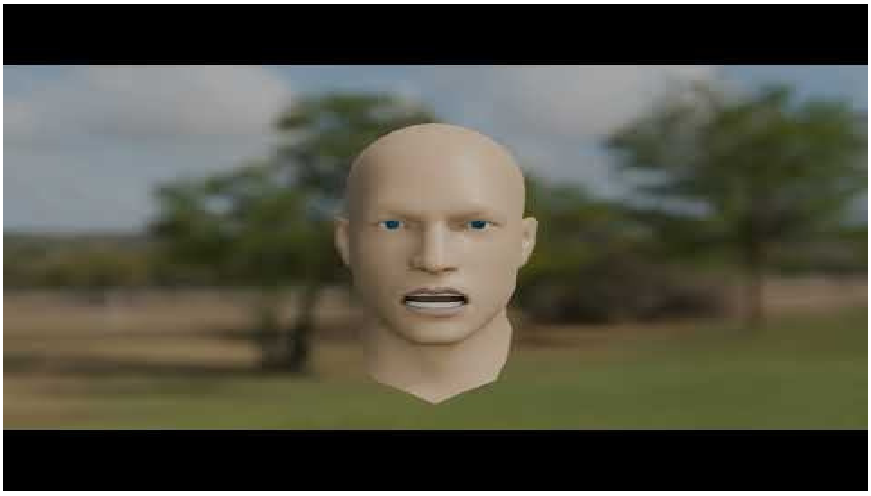
Soru 34



Hangi sesi duyuyorsunuz?

- ma
- va
- ba
- fa
- hiçbiri

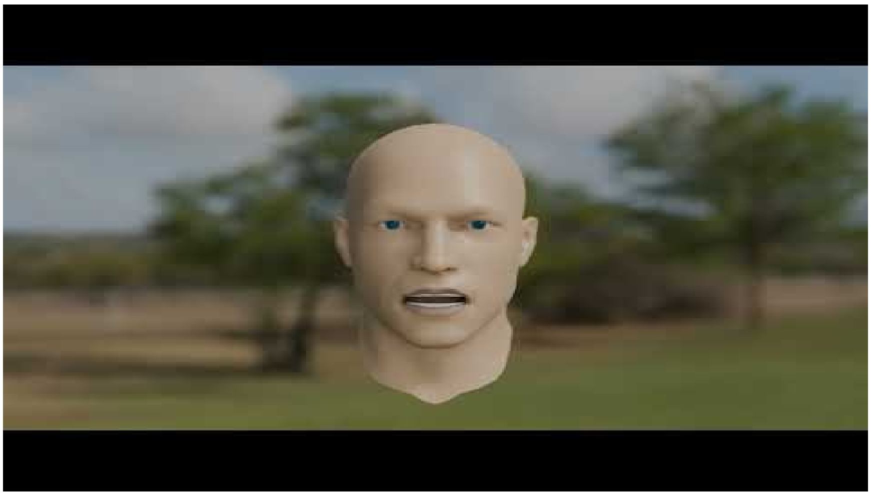
Soru 35



Hangi sesi duyuyorsunuz?

- fa
- ba
- ma
- va
- hiçbiri

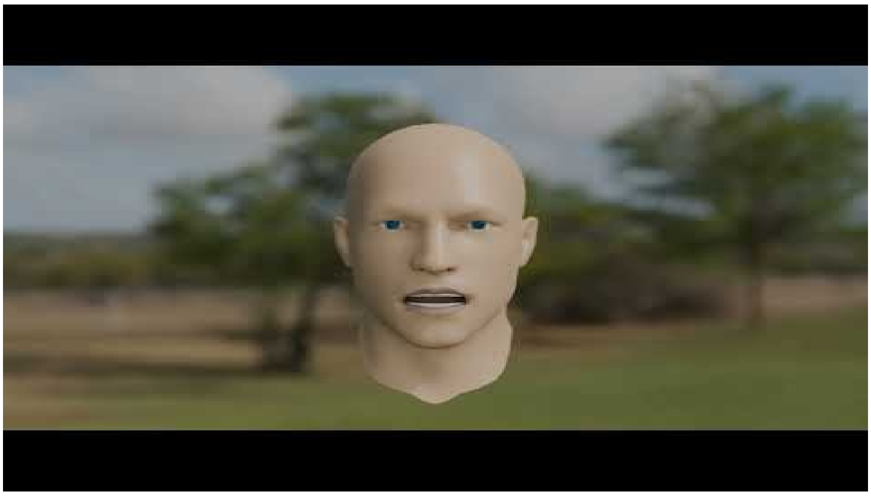
Soru 36



Hangi sesi duyuyorsunuz?

- ba
- va
- fa
- ma
- hiçbiri

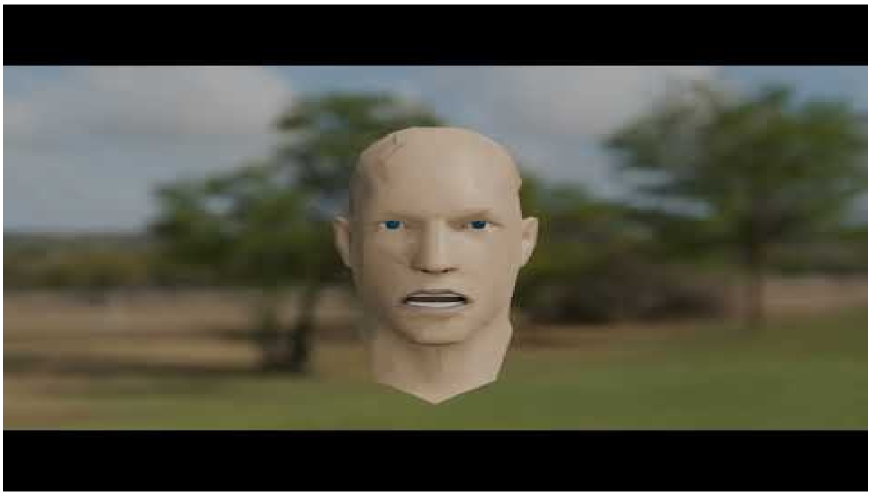
Soru 37



Hangi sesi duyuyorsunuz?

- fa
- va
- ba
- ma
- hiçbiri

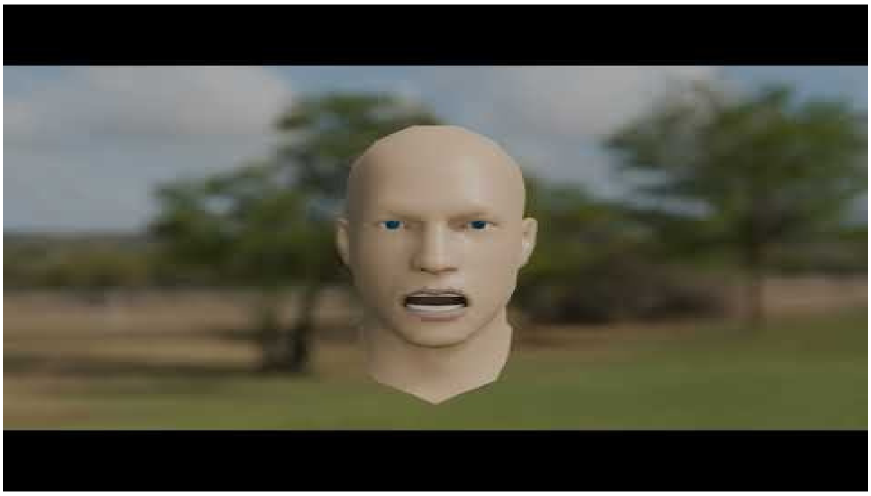
Soru 38



Hangi sesi duyuyorsunuz?

- va
- fa
- ba
- ma
- hiçbiri

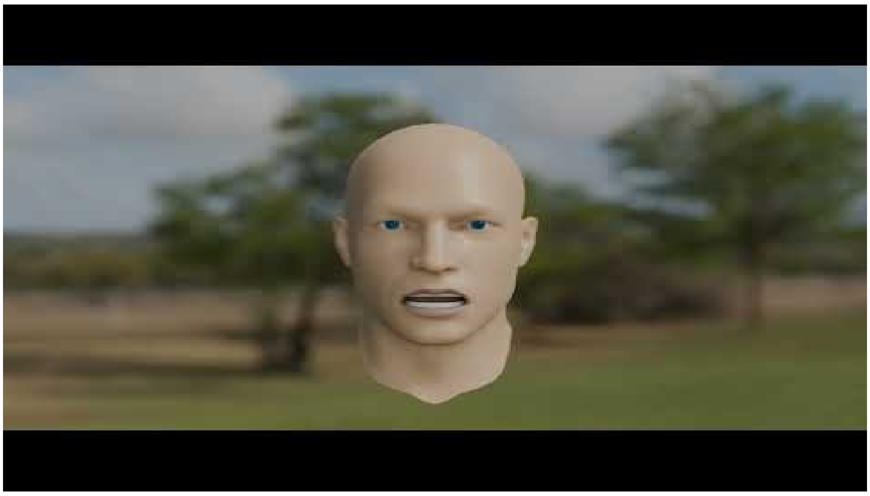
Soru 39



Hangi sesi duyuyorsunuz?

- pa
- ba
- fa
- va
- hiçbiri

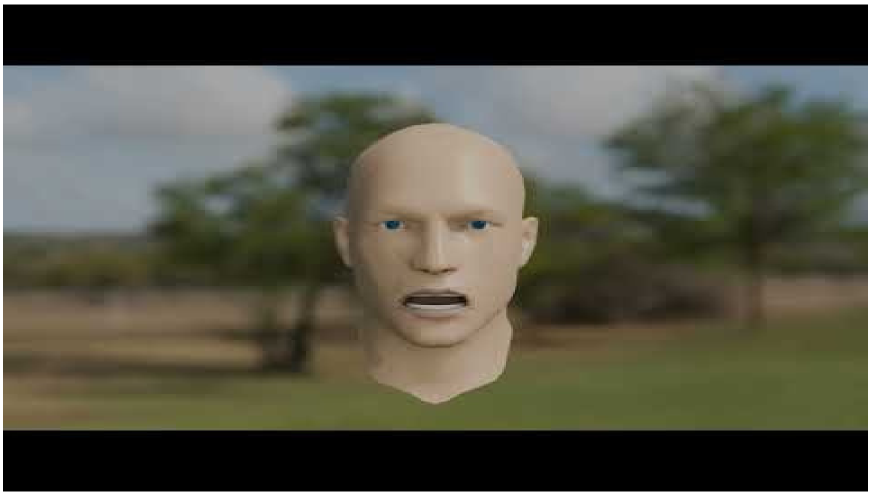
Soru 40



Hangi sesi duyuyorsunuz?

- ma
- va
- fa
- ba
- hiçbiri

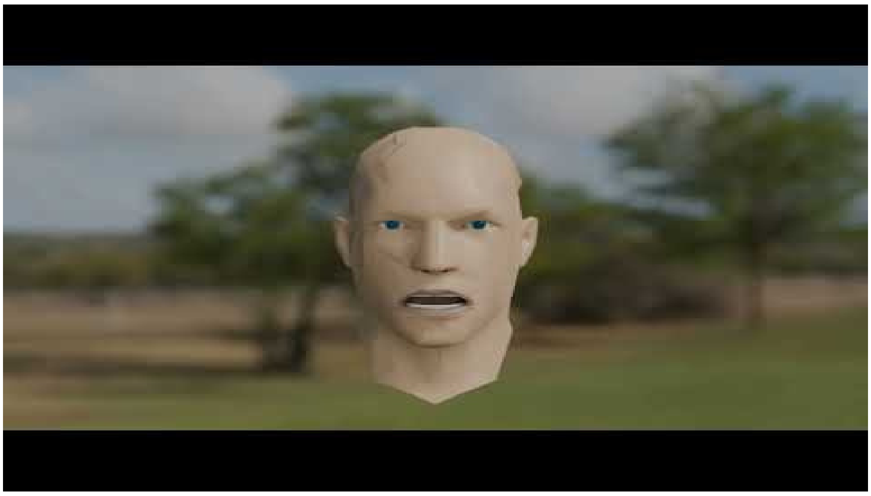
Soru 41



Hangi sesi duyuyorsunuz?

- ba
- va
- ma
- fa
- hiçbiri

Soru 42



Hangi sesi duyuyorsunuz?

- pa
- ba
- fa
- va
- hiçbiri