Revista Infoacceso
e-ISSN 2311-7605 (Online)
infoacceso.org/index.php/ri

Volume 13 · Issue 3 · 2026
Pages 2978–2994

AI-Based Voice Tone Conversion and its Impact on Media Signification: A Semiological Study

Dr. Amina Ferhati

University of Larbi Tebessi – Tébessa, Algeria, amina.ferhati@univ-tebessa.dz

Received: 06/03/2026 | **Accepted:** 13/07/2026 | **Published:** 29/10/2026

***Abstract*—**This study investigates the impact of artificial intelligence (AI)-based voice tone transformation on media meaning through a semiological approach, focusing on how AI-generated vocal tones contribute to meaning production and the construction of credibility within contemporary media discourse. The study adopts a descriptive-analytical design, using the semiological method as its primary framework to examine auditory and visual signs, supported by descriptive and technical approaches to interpret the formal, communicative, and technological characteristics of the selected cases. The research population consists of media models that employ AI technologies to present news content. A purposive sample of three cases was selected: a virtual news anchor used by Al Jazeera Mubasher during the 2024 U.S. presidential election, the BBC's virtual journalist "Aisha Gaban," and a live AI-conducted guest interview broadcast by Al Jazeera Mubasher. The findings indicate that contemporary voice modulation technologies can effectively reproduce key characteristics of professional news delivery, particularly clarity, stability, neutrality, consistency, and calmness. These features reproduce conventional semiotic signs associated with media authority, professionalism, and objectivity, thereby strengthening the perceived credibility of AI-mediated communication. Consequently, AI-generated voices function not merely as technical tools for transmitting information but also as semiotic elements that contribute to constructing media identity, shaping audience perceptions, and producing implicit meanings within news discourse. Despite their increasingly realistic qualities, AI-generated voices remain distinguishable from human voices, particularly in emotional expression, tonal variation, contextual sensitivity, and spontaneous interaction. This distinction reveals a gap between the technological simulation of professional vocal performance and the human dimensions of communication. The study concludes that AI-generated voices represent a significant transformation in the production and perception of media credibility. Their growing integration into journalism raises important ethical, professional, and legal questions concerning transparency, authenticity, accountability, and audience trust. Therefore, the development of appropriate regulatory and professional frameworks is necessary to ensure responsible use of AI technologies while preserving credibility, ethical standards, and public confidence in digital media.

***Keywords*—**Artificial Intelligence, Voice Tone Transformation, Vocal Tone (Tone), Media Meaning, Semiological, Virtual News Anchor, Digital Media, Media Discourse.

1. Introduction

The contemporary media landscape is undergoing rapid transformation through advances in **artificial intelligence (AI)**. Beyond generating textual and visual content, AI can now simulate, clone, and transform the human voice, including its vocal tone. This development has transformed media production while raising important questions concerning the authenticity and meaning of media discourse in the digital environment.

Far from being a mere vehicle for linguistic data, the voice functions as a complex semiotic mechanism, simultaneously transmitting emotional affect, identity markers, and persuasive intent. So when we think about tone we should look at what some important people have said about signs and meanings. For example Saussure said that there are two parts to a sign: the signifier and the signified (De Saussure, 1985). Then there is Peirce who

© 2026 The Author(s). Published by Revista Infoacceso.

said that signs have three parts: the representamen, the object and the interpretant (Peirce, 1931–1958). These ideas help us understand how vocal tone works. Vocal tone is not something that is in the background when we talk. It is actually an important part of the message. It helps us understand what the words mean.

Accordingly changing the sound of a voice using intelligence is not just a technical change; it can alter emotional connotations and perceptions of **authenticity, credibility, authority, and trustworthiness**. In Barthes' perspective, media signs generate both denotative and connotative meanings, meaning that changes in vocal tone may reconstruct the implicit meanings of media messages and influence audience interpretation (Barthes, Image, music, text , 1977).

The development of AI-based voice transformation technologies therefore creates new opportunities for adapting media messages to different audiences while simultaneously raising concerns about media credibility and the manipulation of vocal signs. Consistent with McLuhan's view that the medium can reshape the message and its perception, AI may be understood as a new medium capable of restructuring aspects of media meaning (McLuhan, 1964).

Despite increasing research on AI, voice cloning, and voice transformation, studies examining the **vocal sign from a semiological perspective** remain limited. This study addresses this gap by examining how AI-based voice tone transformation affects meaning production, interpretation, and audience reception within media discourse.

1.1 Research Problem

The study seeks to answer the following central question:

To what extent does AI-based voice tone transformation influence the construction of media meaning from a semiological perspective, and how does this transformation reconstruct the vocal sign, guide interpretation, and shape meaning for audiences?

1.2 Research Questions

- How is vocal tone constituted as a semiological sign within media discourse?
- How does AI reconstruct the vocal sign through voice tone transformation?
- What denotative and connotative transformations result from voice tone transformation?
- How does transformed vocal tone affect meanings of credibility, emotion, and identity?
- What does semiological analysis reveal about its influence on meaning construction and media interpretation?

1.3 Significance of the Study

This research is about where artificial intelligence and media studies meet. It looks at how people communicate in the world. The main idea is that the way someone sounds when they talk is not a side thing it actually means something. So this study looks at how fake or changed voices affect how people feel how they listen and if they think something is true or not.

The point of this work is important in two ways. First it helps fix a problem in what people're studying now by using old ideas about signs and symbols to look at how artificial intelligence changes voices. Second by understanding how these fake voices work we can make rules and guidelines for media, which is something we really need to do. Artificial intelligence and digital media are closely related to this research. It is all, about artificial intelligence and how it affects our communication.

1.4 Objectives of the Study

The study aims to:

- Examine the impact of AI-based voice tone transformation on media meaning.
- Analyze vocal tone as a semiological sign.
- Identify differences between natural and AI-generated vocal tone.
- Examine how AI reconstructs the vocal sign and its relationship to the signifier and signified.
- Identify semantic transformations resulting from voice tone transformation.
- Analyze its influence on credibility, emotion, identity, and persuasion.
- Examine its role in audience interpretation and meaning construction.

1.5 Research Methodology

The study adopts the **descriptive approach** to examine AI-based voice tone transformation and establish its theoretical and media-related dimensions. For the empirical analysis, it employs the **semiological method**,

treating vocal tone as a meaning-producing auditory sign and examining its emotional, persuasive, and cultural dimensions. (Widdowson, 1996)

The semiological analysis examines the relationship between the vocal **signifier and signified**, as well as denotative and connotative meanings and the interaction between vocal tone and verbal and non-verbal elements. The combination of descriptive and semiological approaches therefore allows both the phenomenon and its underlying meanings to be examined.

1.6 Scope of the Study

The study is limited to the impact of **AI-based voice tone transformation on media meaning**, focusing on vocal tone as a semiological sign. Methodologically, it combines descriptive and semiological approaches. Spatially, it examines media content published through digital platforms and media organizations using AI-based voice technologies. Temporally, it focuses on the period **2022–2026**, corresponding to significant developments in generative AI for voice synthesis and transformation.

2. Theoretical and Conceptual Framework of AI-Based Voice Tone Transformation

2.1 The Concept of Voice Tone Transformation and Its Technologies

2.1.1 The Concept of Voice Tone

The study of phonetics looks at how speech sounds work. It includes how we make these sounds what they sound like and how they help us communicate. When we talk the tone of our voice is a part of how we sound. It is not something that goes along with what we say it actually helps us understand what we mean what we want and how we feel. The tone of our voice works with things like how we raise and lower our voice, which words we stress and the rhythm of our speech to make sure people understand us.

This way of thinking is similar to what Nikolai Trubetzkoy said about phonetics. Nikolai Trubetzkoy looked at speech sounds in a physical way. He said that things like stress and how sound waves behave and how they move from the person talking to the person listening are all important, to understanding how speech sounds work. Nikolai Trubetzkoy thought that if we want to understand phonetics we have to think about the physical parts of speech sounds. (Iluwar, 1980)

Stress constitutes an important phonological feature in Arabic and many other languages. It is produced through greater articulatory force, which gives prominence to a particular sound or syllable (Faris, 1991). A stressed syllable consequently requires greater muscular effort than neighboring syllables, making it more prominent and perceptible to the listener (Tamam, 1980).

Thus, stress can be understood as the prominence given to a particular syllable through increased articulatory force. It contributes to the organization of the sound structure of words and may, in certain languages, perform a semantic function by distinguishing meanings.

2.1.2 Artificial Intelligence Technologies Used in Voice Transformation

Speech processing has witnessed significant development due to advances in artificial intelligence (AI), particularly deep learning. Unlike conventional speech-processing systems based mainly on rule-based programming and limited statistical models, contemporary systems employ deep neural networks capable of learning linguistic and acoustic patterns from large datasets. This progress has improved automatic speech recognition, enhanced text-to-speech synthesis, and enabled the generation of synthetic voices that increasingly approximate natural human speech (Goodfellow, 2016).

Modern voice transformation applications rely on several specialized AI models. Among the most influential is WaveNet, developed by DeepMind, which generates natural speech waveforms using causal convolutional neural networks and represented a significant improvement in the quality of synthetic speech (Van den Oord, 2016).

Another important development is Tacotron, which uses an attention mechanism to convert text into speech representations, followed by Tacotron 2, which further improved the naturalness, clarity, and intelligibility of synthesized speech (Wang, 2017).

In Automatic Speech Recognition (ASR), Whisper, developed by OpenAI, has demonstrated strong multilingual speech recognition capabilities and supports both transcription and speech translation across diverse acoustic conditions (Radford, 2022).

More recently, VITS has integrated text-to-speech synthesis and speech generation within an end-to-end learning architecture, contributing to improvements in the speed and quality of speech synthesis (Kim, 2021).

These improvements create the building blocks for many modern audio uses. Of working alone the parts that make smart helpers, instant translation and fake voice copying work now come together to change how people send and receive digital messages. This mixing is important in areas like online learning and making things easier for people with disabilities, where advanced ways to turn text into speech are becoming more important. Because of this the quick growth of these computer systems shows more, than small improvements in how well they work or how clear the sound is. It means there is a big change happening making voice changing tools a part of todays media world.

2.1.3 The Evolution of Voice Cloning and Voice Transformation Technologies

Among the technologies most relevant to the present study are Deepfake technology and Voice Cloning, both of which have contributed significantly to the development of AI-based voice and audiovisual transformation.

Deepfake Technology

Fundamentally, deepfakes constitute a sophisticated class of audiovisual manipulation driven by artificial intelligence. By synthesizing highly convincing replicas of a subject's vocal patterns, facial geometry, or physical gestures, this technology successfully projects the illusion of events or statements that never took place in reality (Silverman, 2020).

At their core, these architectures rely heavily upon deep learning protocols and neural networks processed through massive datasets. Because these kinds of frameworks carefully recreate facial expressions the way people speak and how lips move the fake videos made this way always make it hard to tell what is real. Because of this it becomes really hard for normal people to notice when something is not real (Alanaz, 2024).

Awareness about this problem started in 2017 when a person on Reddit who went by the name "Deepfake" shared videos that had faces of famous people. These videos used computer programs to make the faces look real. After this time people saw it the technology spread very fast all over the internet. It got used in ways for businesses but also caused big problems with the law and big questions, about right and wrong (Al-Qadri, 2024).

Voice Cloning Technology

Voice Cloning refers to technology capable of generating speech that closely resembles the voice of a target individual and allows the system to produce arbitrary words and phrases from a limited number of voice samples. Through deep learning algorithms, it can reproduce important characteristics of the original speaker, including tone, pitch, and speaking style (Kadam, 2024).

The combination of learning and advanced neural networks has really changed how we make voice clones sound. Now we can do more than just get the basic tone right. We can make voice clones that sound like they are, from a place and have the same pitch and rhythm as a real person. This means we can make voices that are very hard to tell apart from real voices (Gerard, 2023).

The fact that we can now create videos and audio that appear real even though they are not has changed the way we create content. We can make it seem like a person is saying or doing something even if they are not. This is important because it shows that computers can imitate not how someone looks but also how they sound. Voice cloning and Deepfake are coming together to make this possible.

2.2 Semiology and Meaning in Media Discourse

2.2.1 Concept and Major Theorists

Operating as the comprehensive study of symbolic frameworks, semiology interrogates exactly how societies construct, negotiate, and exchange meaning. This discipline is anchored in the understanding that human communication fundamentally transcends traditional spoken or written linguistics. The way we understand things is not about what we hear or see. It is about all the things around us that help us make sense of the world. We get meaning from pictures and sounds and also from the way people move their bodies. Each of these things helps us understand something in its way depending on where we are and what is acceptable, in that place. When we look really closely at how we figure out what things mean we can see how it all works. We can see how the thing that is being shown to us and the idea it is supposed to represent are connected. This helps us understand how people figure out what complicated messages are saying, no matter where they see or hear them (De Saussure, 1985).

Revised Academic Text

The genesis of modern semiology is deeply anchored in the theories of Swiss linguist Ferdinand de Saussure, who originally envisioned a comprehensive science mapping the societal life of signs. Central to Saussure's framework is the indivisible, dyadic bond between the signifier (the acoustic form) and the signified (the mental abstraction). Crucially, he posited that this structural linkage remains entirely arbitrary, sustained purely by collective social convention (De Saussure, 1985). Conversely, american philosopher Charles Sanders Peirce constructed a more expansive, triadic architecture. By dividing meaning-making into the representamen, the object, and the interpretant, Peirce's paradigm explicitly pushed semiotic inquiry beyond verbal constraints, establishing a robust framework for analyzing a vast array of non-linguistic communicative phenomena (Peirce, 1931–1958).

The study of language and culture has. So has semiology. Now semiology is a way to look at things like books, media, movies and ads. Roland Barthes was a person in making this happen. He used semiology to find meanings and cultural myths in things we read and see. Semiologists like Roland Barthes use semiology to understand semiology in our lives. The semiology that Roland Barthes worked on is still important, for semiology today.

He emphasized that signs communicate not only literal meanings but also culturally and ideologically constructed significations (Barthes, Elements of semiology , 1967).

In media analysis, semiology examines journalistic and communicative messages as systems of signs, including linguistic, iconic, and kinetic forms. Its scope therefore encompasses words, speech sounds, gestures, symbols, and the relationships between communicative and behavioral phenomena and systems of meaning (Widdowson, 1996).

In the digital era, semiological analysis has also become relevant to artificial intelligence-generated content, including synthetic images, cloned voices, and deepfake videos. Such technologies produce visual and auditory signs capable of simulating reality with a high degree of fidelity. Consequently, semiology provides an appropriate theoretical framework for examining their structures of meaning, mechanisms of meaning production, and potential influence on audience perception and interpretation.

Vocal Meaning and Its Components

Vocal meaning refers to the semantic contribution of speech sounds to the expression of meaning within words and utterances. It is shaped by the phonetic composition of words, particularly their consonants and vowels, as well as by sentence structure and patterns of oral delivery (Okasha, 2005).

Vocal meaning may be classified into two categories. The first category, systematic functional meaning, operates strictly within established phonological parameters; it arises directly from the structural arrangement or substitution of individual phonemes within a lexical unit. Conversely, non-systematic vocal meaning functions independently of these rigid rules, deriving instead from suprasegmental phenomena. This secondary layer of meaning relies entirely on performative prosodic features—such as dynamic stress, intonation contours, and strategic pausing. Consequently, it shapes the interpretive nuance of live speech delivery without ever altering a word's foundational phonological architecture (Mujahid, 1985).

Within linguistic and semiotic studies, one of the principal components of vocal meaning is stress. Stress refers to the prominence assigned to a particular syllable within a word or sentence through increased articulatory force, making it more perceptible than surrounding syllables. According to Ibrahim Anis, proper stress placement is essential to accurate pronunciation, as differences in stress patterns may make speech sound foreign and may contribute to ambiguity in meaning. Kamal Bishr further distinguishes between stress languages, such as English, German, and Spanish, and non-stress languages, such as Japanese and Arabic, while differentiating between stress and accent (Anis, 1999).

2.2.2 Intonation

Intonation is about how the pitch of our voice changes when we talk. It helps us express things like questions, excitement, importance and instructions. It also helps the person listening understand what we mean. When we change the pitch of our voice it can change the meaning of what we're saying. Intonation can also tell people something about who we're how we are communicating with each other like where we come from and what we are talking about (Bouchair, Phonetics, 2000).

2.2.3 Rhythm

Rhythm is, like the beat of our speech. It is the way we say syllables that makes our speech sound smooth and easy to listen to. Our brains like to hear rhythm in speech because it helps us understand what the other person is saying. When we talk we do not just say things in a voice. We change the pitch of our voice when we say syllables. Sometimes the way we change the pitch can actually change the meaning of what we're saying (Anis, 1999).

2.2.4 Pitch

Pitch is how high or low our voice sounds. It is decided by how our vocal cords vibrate. Pitch helps us express our emotions when we talk. In some languages pitch is also used to distinguish between meanings of words.

2.2.5 Loudness (Intensity)

Loudness, or **intensity**, is how strong or weak our voice is when we talk. We use loudness to stress words or to show how we are feeling like if we are angry, happy or excited.

2.2.6 Duration

Duration denotes the amount of time required to articulate a sound or syllable. In certain languages, duration performs a semantic function while also influencing speech rhythm and intelligibility.

2.2.7 Pauses

Pauses are brief intervals of silence occurring during speech. They contribute to discourse organization by marking semantic units, emphasizing meaning, and providing listeners with opportunities to process and interpret the message.

2.2.8 Voice Quality (Timbre)

Voice quality, or **timbre**, encompasses a speaker's unique acoustic signature, spanning from hoarse and breathy to exceptionally clear. Beyond acoustics, these textured traits actively transmit socio-emotional meaning, operating as calculated persuasive mechanisms within media discourse to shape audience perception (Bouchair, Phonetics, 2000).

Media Language and the Role of Vocal Performance in Meaning Construction

When we use Modern Standard Arabic in a way it is very good for talking to a lot of people. This language is easy to understand. It can convince people. The way we speak is very important because it helps people understand what we are saying (Belaid).

For example when we're, on the radio or television the way we say things is crucial. We use rhythm. We pause on purpose and our tone of voice helps people understand what we mean. Modern Standard Arabic is used in a way that helps people believe what we are saying. When we are speaking on the radio or television we might change the way we speak to fit the situation. We still use the basic rules of Modern Standard Arabic (Willayid, 2003).

These variations may produce pronunciation shortcomings that are particularly noticeable in broadcast media. The phonetic dimension of media language includes pronunciation and vocal performance, lexical aspects, and grammatical aspects, which are interconnected in spoken communication (Omar, 1993).

According to Said Boukhaouch, such phonetic variations are particularly noticeable in television news and journalistic interviews, where broadcasters' voices can either reinforce correct linguistic usage or contribute to the spread of incorrect pronunciation patterns (Boukhaouch, 2019).

3. **The Impact of Voice Tone Transformation on Media Meaning**

3.1 Section One: Semantic Transformations Resulting from Voice Tone Modification

The Impact of Voice Tone Changes on the Reception of Media Messages

When we watch the news or listen to someone on the radio the way they speak is really important. The media people use their voice to help us understand what they are saying and to make us like the place they work for. They choose their words carefully. Try to sound nice so we keep listening. If they sound friendly we feel like we know them. We want to keep watching.. If they sound serious and professional we think they are smart and know what they are talking about especially when they are reading the news (Biermann, 2025).

The people on the radio and tv also need to be good at speaking and using their voice in different ways. Some of them do not get training on how to say words correctly and use their voice well. This means they cannot use all

the sounds they could be using. If they know how to use their voice they can make their words sound more interesting by being louder or softer, faster or slower and by pausing at the times. These things help us understand what they are saying and make them sound more believable. The media people need to be good, at using their voice so we can understand what they are saying and think they are trustworthy (Omar, 1993).

The Relationship Between Voice, Credibility, and Persuasion

The philosopher **Friedrich Nietzsche** succinctly captured the relationship between vocal tone and persuasion when he observed: *"We often reject something simply because the tone in which it is expressed fails to engage us."* This profound insight argues that intellectual acceptance relies heavily upon a broadcaster's physiological delivery, rather than operating solely on textual merit. Consequently, human vocalization transmits critical subtextual data parallel to spoken language, operating synergistically to either actively secure audience alignment or inadvertently trigger cognitive resistance.

Parallel acoustic investigations confirm that modulated delivery actively drives cognitive focus toward targeted subjects, effectively diffusing psychological friction while establishing affective equilibrium. Therefore, maintaining a stabilized, deliberate, emotionally regulated cadence consistently maximizes persuasive impact throughout any complex public media discourse.

Supplementary behavioral analyses reveal that suprasegmental execution broadcasts vital metrics regarding internal psychological conditions, telegraphing affective states ranging from severe volatility to profound composure. Acoustic delivery fundamentally bypasses literal linguistic boundaries, acting as a transparent barometer for hidden intentionality. This paradigm mirrors the Qur'anic assertion regarding deceptive individuals: **"You will surely recognize them by the tone of their speech."** Ultimately, the precise physiological execution of speech actively exposes concealed motives, continuously transmitting deeper truths that exist entirely outside the vocabulary itself, thereby reshaping the ultimate media message interpretation completely (Ashour).

Psychological and Communicative Effects on Audiences

Vocal intonation operates as a primary catalyst for triggering affective reactions within media consumers. This mechanism fundamentally transcends basic data transmission, actively directing audience sentiment while simultaneously dictating their cognitive and behavioral outcomes. Crucially, this psychological manipulation fluctuates heavily depending upon specific acoustic parameters, namely precise pitch, structural rhythm, and the profound emotional resonance projected.

Deploying a measured, exceptionally warm cadence rapidly establishes institutional trust, effectively cultivating an atmosphere characterized by profound psychological security and deep cognitive reassurance. Strategically, broadcasters frequently direct this specific acoustic profile toward demographics experiencing high uncertainty or persistent anxiety. Delivering information gently constructs a highly stable affective environment, implicitly prompting these hesitant listeners to embrace the underlying narrative and react favorably overall.

Conversely, executing a highly energetic and aggressively enthusiastic vocal register effectively injects immediate motivation, strictly engineered to force rapid consumer mobilization. This precise tactic dominates modern commercial advertising and targeted public awareness operations, where professional audio performers deploy intensely dynamic expressive ranges to optimize physiological arousal. Ultimately, this intense acoustic momentum aggressively pushes listeners toward decisive material engagement, triggering immediate product acquisitions, deliberate social participation, or broader campaign compliance (Soundeals.).

3.2 A Semiological Analysis of Artificial Intelligence Applications in Media

Analysis of Media Models Employing Voice Tone Transformation

This section examines a selection of media models presented by various broadcasting organizations that employ artificial intelligence technologies in the delivery of news content. The analysis focuses on describing these models through an examination of their vocal tone characteristics and performance styles from a semiological perspective.

It is important to note that some of these models were produced entirely using artificial intelligence, featuring virtual news anchors with no real human counterpart, whereas others represent different forms of AI-assisted media production in which artificial intelligence supports or enhances journalistic presentation.

Model One

Title: *"Al Jazeera Mubasher's AI News Anchor Launches Special Coverage of the 2024 U.S. Presidential Election"*

The first model includes a 54-second opening news part. In this part Al Jazeera Mubasher used an AI-created news presenter to introduce its special coverage of the 2024 United States presidential election. This model gives an example for looking at how AI-created voice style affects the making of media meaning, professional trustworthiness and what people think about the news, in todays news broadcasting (Mubasher., AI presenter on Al Jazeera Mubasher launches special coverage of the 2024 U.S. presidential election, 2024).



Extracted Image from the Video

First: Descriptive Approach

The video adopts an introductory format characteristic of major news coverage openings, aiming to capture viewers' attention while emphasizing the significance of the event. Its principal features include:

- The appearance of a **digital news anchor** with highly realistic facial features and limited yet well-coordinated facial movements.
- The use of **clear and direct Modern Standard Arabic**.
- A presentation that frames the U.S. presidential election as a global event of considerable significance before transitioning seamlessly into the news coverage.
- A modern visual design employing graphics and colors consistent with **Al Jazeera Mubasher's** visual identity, together with a fast-paced rhythm appropriate for live broadcasting.

Analysis of the Vocal Tone

From an auditory perspective, the vocal performance can be characterized as follows:

a. High Degree of Formality

- The vocal tone closely replicates the delivery style of conventional television news anchors.
- It exhibits little evidence of personal or emotional expression.

b. Confidence and Composure

- The volume remains relatively consistent throughout the presentation.
- There are no abrupt fluctuations in vocal intensity or speech rate.

c. Moderate-to-Fast Delivery

- The speaking rate is appropriate for a news introduction.
- The pacing conveys energy without appearing rushed.

d. Clear Articulation

- Pronunciation is exceptionally clear, with careful attention paid to articulation.
- The performance contains minimal hesitation or pronunciation errors typically associated with spontaneous human speech.

e. Limited Intonation

- Minor pitch variations occur only when emphasizing key words.
- Compared with a professional human news presenter, the intonation remains less varied, creating the impression of an AI-generated or AI-modified voice.

Overall Auditory Impression

The voice does a job of achieving what a news introduction is supposed to do. It gives information clearly. Sounds confident and professional.. Some people might find it less friendly or less lively, than a real person. This is because:

- Its highly regular rhythm;
- The limited occurrence of natural pauses; and
- The restricted degree of emotional expression throughout the sentences.

Second: Technical Approach

This model represents an advanced application of **generative artificial intelligence** within the media industry, employing a virtual news anchor to present the opening segment of the broadcaster's coverage of the **2024 U.S. presidential election**. Technical analysis reveals the integration of several AI technologies, including facial image generation, speech synthesis, and automated lip synchronization, all designed to replicate human presentation with a high degree of realism.

Visual Level

The virtual presenter exhibits highly realistic facial features while maintaining a visual identity consistent with the broadcaster's news-oriented brand. Head movements and facial expressions are deliberately restrained and carefully controlled, reflecting the formal conventions of television news broadcasting. However, they do not display the expressive diversity characteristic of natural human performance. The synchronization between lip movements and spoken content is highly accurate, indicating the use of advanced facial animation technologies driven by speech signals.

Vocal Level

The speech is distinguished by clear articulation, consistent delivery speed, and stable vocal intensity throughout the presentation. The vocal tone remains neutral and composed, characteristics that conform to established professional standards for news broadcasting. Nevertheless, the performance demonstrates limited variation in intonation and relatively restrained emotional expression compared with a human presenter. Transitions between sentences are highly regular, while the subtle pitch fluctuations, hesitations, and natural vocal irregularities typically associated with human speech are largely absent.

Technological Level

The model appears to rely on advanced **Text-to-Speech (TTS)** technology supported by deep learning architectures, together with **AI avatar generation** and **lip synchronization (Lip Sync)** technologies. These systems enable the production of a virtual news presenter capable of delivering news content coherently and with a high degree of technical quality.

The overall video production also preserves the visual identity of **Al Jazeera Mubasher** through the use of consistent graphics, newsroom backgrounds, and stable camera angles. These production choices reduce the perceived distinction between the virtual presenter and a human broadcaster, thereby enhancing the credibility of the viewing experience.

Overall, this model demonstrates a highly advanced level of AI-driven media production by achieving strong synchronization between visual and auditory components while preserving the professional characteristics expected of television news delivery. Nonetheless, the limited diversity of facial expressions and vocal intonation remains one of the clearest indicators of the presenter's artificial nature, despite the remarkable technological sophistication achieved.

Third: Semiological Approach

The sound of a news presenters voice really matters when it comes to understanding and perceiving the news. A news presenter usually speaks in an consistent way with a clear rhythm and tone which makes them sound calm and trustworthy. This is what people expect from news, where the voice is a way to get information across while sounding fair and knowledgeable.

Over time the way news presenters speak has changed from a natural way of talking to a more artificial and polished way. Unlike how people talk, with ups and downs in their voice and emotions a news anchors voice stays calm and neutral without much personality or feeling. This helps people think the news is believable and important.

The way news is presented is smooth and straightforward without the ups and downs of speech. This makes the news sound serious and factual. Sometimes it can sound a bit fake. The news is presented in an precise way without interruptions, which is what people expect from a professional news presenter.

The way news presenters sound has changed a lot from sounding like a person to sounding like an institution. This means that a news presenters own voice is replaced by a way of speaking that is created by computers. This shows that there is a move towards an idea of what a news presenter should sound like, where the institution is more important than the individual.

The way news presenters speak sends a messages. It says that the news is neutral, professional and trustworthy. It also shows that it is not a natural way of speaking. This is part of a change, in the way media uses voice from having a unique person presenting the news to having a computer-generated voice that follows professional rules and represents the institution. This is changing the way we get our news and information.

Model Two

Title: *"Virtual Journalist Aisha Gaban"* – **BBC**

The second model examines the virtual journalist **Aisha Gaban**, introduced by the **BBC** during a television program discussing the impact of artificial intelligence on the future of the labor market. In this program, the broadcaster employed an AI-generated virtual journalist to present segments of the discussion, providing a practical example of how artificial intelligence can be integrated into contemporary news production and presentation (Youm.).



Extracted Image from the Video

First: Descriptive Approach

This documentary, entitled **"Will Artificial Intelligence Take My Job?"**, explores the transformations that artificial intelligence is expected to bring to the labor market across various sectors, particularly music, fashion, medicine, and law (News.).

The report adopts a continuous narrative style supported by explanatory footage and visual graphics that enhance the audience's understanding of the topic. At the conclusion of the documentary, the presenter reveals to

viewers that she is not a real person but rather a character generated entirely through artificial intelligence technologies, with both her image and voice digitally created.

The video aims to familiarize audiences with the remarkable progress achieved by artificial intelligence technologies and to demonstrate their ability to simulate human presenters in media production. It also seeks to raise public awareness that digital content is becoming increasingly difficult to distinguish from AI-generated content, thereby emphasizing the growing importance of verifying information sources.

The descriptive analysis indicates that the video employs integrated visual and auditory elements to create a highly realistic media experience, in which the virtual journalist **Aisha Gaban** appears indistinguishable from a human news presenter. The story uses a news style. It moves between parts. It talks about intelligence as the main topic. It also uses intelligence to make the story. At the end people find out the person speaking is not real. It is a computer made model. This is meant to show how good current artificial intelligence is. It also wants to start a conversation, about what this means for news and jobs.

Second: Technical Approach

The video relies on **generative artificial intelligence** technologies to create a virtual media personality represented by the journalist **Aisha Gaban**. This process combines several digital technologies, including virtual human generation, facial expression simulation, and AI-generated speech synthesis designed to closely resemble natural human speech, with the objective of producing a television presenter who appears convincingly realistic.

Artificial intelligence therefore serves not merely as the subject of the report but also as the principal tool employed in the production of the media content itself. Algorithms are utilized to generate a virtual personality capable of presenting an entire television program independently.

The people at **Seraphinne Vallora** made a computer presenter. This company likes to use intelligence for fashion. They did this work for **Kalel Productions**. The team gave the computer lots of instructions to make it seem real. They wanted the computer to talk and act like a person on camera. The computer presenter is so good that you can not tell it is not a person until the end of the documentary. That is when the people who made the documentary tell the viewers the truth, about the presenter (Channel., 2025).

Visual Technologies Employed

The video clearly demonstrates the use of advanced digital image-processing technologies, including:

- Virtual character modeling: creating a digital personality with realistic human facial features that closely resembles a traditional television journalist.
- Facial animation: synchronizing lip movements and facial expressions with the spoken dialogue.
- Human visual simulation: employing AI techniques to make the virtual character appear natural through realistic eye movements and gestures.
- Digital editing: integrating the virtual presenter with supporting visual materials illustrating the impact of artificial intelligence across various professional sectors .

Audio Technologies Employed

The video utilizes several advanced speech-processing technologies, including:

- AI-generated synthetic speech that imitates the delivery style of professional television presenters.
- Careful adjustment of speech rhythm and vocal tone to suit the conventions of news reporting.
- Precise synchronization of the synthesized voice with the movements of the virtual presenter to achieve audiovisual coherence.

Together, these technologies enhance the realism of the virtual journalist, enabling her to perform the complete communicative functions of a television presenter.

Production and Editing Techniques

From a production perspective, the video employs techniques commonly found in contemporary television reports, including:

- Seamless transitions between the virtual presenter and supporting visual footage.
- The use of graphics and explanatory visuals to simplify the discussion of artificial intelligence.
- A narrative sequence organized around an introduction, illustrative examples, and a concluding revelation of the presenter's virtual identity.

- A relatively fast editing pace consistent with the conventions of digital media content.

Third: Semiological Approach

Voice as a Semiological Sign

In this video, voice is employed not merely as a technical component but as a **semiological sign** that contributes to constructing the identity of the virtual journalist while endowing her with a presence closely resembling that of a human presenter. Voice constitutes one of the principal cues through which audiences evaluate the nature of a speaker, being culturally associated with trustworthiness, authority, credibility, and professional competence.

The voice selected for **Aisha Gaban** deliberately imitates the delivery style of traditional television news presenters through clear articulation, regular rhythm, and controlled vocal modulation, thereby providing the virtual journalist with a convincing professional media identity.

The Semiological Significance of the News Tone

Aisha Gaban employs a vocal tone closely resembling that of professional news anchors, characterized by several distinctive features.

a. Calm and Stable Tone

Throughout most of the documentary, the voice remains calm, balanced, and composed. Semiologically, this tone signifies:

- trustworthiness;
- objectivity;
- control over the information being presented; and
- the professional character traditionally associated with television journalism.

These characteristics encourage viewers to perceive the virtual journalist as an experienced human broadcaster.

b. Regular Speech Rhythm

The presentation goes at a pace with a steady flow of words. This way of speaking is often used when people report the news. It helps people understand what is being said without getting confused. From a semiological perspective, this rhythm signifies:

- order;
- rationality; and
- the informative nature of the message.

The voice changes a lot when the documentary talks about the future of jobs and how artificial intelligence will affect society. The voice becomes really calm and thoughtful when it discusses these things. It is not just giving us information it is also helping us think about what's going to happen with new technology. The voice is, like a guide that helps us think about the future and what artificial intelligence will mean for us. The documentary is trying to get us to think about intelligence and the voice is a big part of that.

Based on the foregoing analysis, the semiological examination of the voice in **Aisha Gaban's** documentary demonstrates that vocal tone performs a fundamental role in constructing the meaning of the virtual journalist. The voice has been deliberately designed to reproduce the defining characteristics of conventional broadcast speech—clarity, composure, confidence, and organization—while shifts in vocal tone from news delivery to explanation and finally to reflection shape the audience's interpretation of the message.

Perhaps the most significant semiological dimension lies in the paradox of a voice that sounds unmistakably human while originating from an artificial system. In this sense, the voice itself becomes a sign of the technological transformation that is redefining the relationship between human beings and intelligent machines within the contemporary media landscape.

Model Three

Analysis of the Video Entitled "AI News Anchor Replaces an Al Jazeera Mubasher Presenter Live on Air and Interviews a Guest"

The AI-generated virtual presenter appeared on **Al Jazeera Mubasher** on **10 March 2023**, marking the **first Arab experiment** in which an artificial intelligence system replaced a human television presenter during a live broadcast. In this pioneering demonstration, the AI anchor assumed the role of **Maisoon Al-Dakhil**, presenter of the program **Hashtag**, and conducted a live interview

with **Dr. Tab Al-Ashqar** (Mubasher., AI presenter replaces).



Extracted Image from the Video

First: Descriptive Approach

The descriptive approach analyzes this media event by examining its visual and auditory components without making evaluative judgments.

Visual and Aesthetic Components

- **Visual Identity:** The virtual presenter appears as a young Arab woman wearing contemporary formal attire consistent with the cultural and visual identity of **Al Jazeera Mubasher**.
- **Body Language:** The movements of the head and eyes are designed to follow both the camera and the guest, while facial expressions simulate natural human gestures during conversation, thereby reducing the viewer's perceptual gap.
- **Lip Synchronization:** The video demonstrates a medium to high level of accuracy in synchronizing lip movements with spoken phonemes, although a slight mechanical rigidity remains evident, distinguishing the digital avatar from a human presenter.

Vocal and Linguistic Components

- **Voice Tone:** The presenter uses a voice that is made by a computer it sounds like a news anchor on television. The presenter talks at a speed but it does not sound completely natural like a real person talking.
- **Linguistic Accuracy:** The presenter says the words correctly especially when it comes to Modern Standard Arabic so we can understand the news. However the presenter puts much stress on the last part of the words when they are ending with a vowel sound.
- **Intonation:** The computer voice does not sound very emotional it just talks in a way and it does not really react to what the guest is saying it just keeps talking in the same tone, like Vocal and Linguistic Components of the presenter the Vocal and Linguistic Components of the news do not have a lot of feeling.

Second: Technical Approach

AI news anchors are virtual personalities created using advanced machine-learning algorithms. These algorithms analyze massive volumes of data—including news articles, videos, and audio recordings—to generate highly realistic representations of television news presenters.

This technical approach may be analyzed through the following dimensions:

Image Generation and Motion Technologies

AI image generation is one of the uses of generative artificial intelligence. It changes text or data into pictures. These tools use learning models and large language models to understand what people write and connect words with the right images. This leads to creative results. AI image generation is one of the uses of generative artificial intelligence. It changes text or data into pictures. These tools use learning models and large language models to understand what people write and connect words, with the right images. This leads to creative results (Chen, 2025).

Prompt Engineering and Text-to-Speech Technology

The system utilizes advanced prompt engineering and Text-to-Speech technologies to achieve:

- the generation of relatively natural-sounding Arabic speech;
- precise control over speech rate, vocal tone, and rhythm.

Third: Semiological Approach

From a semiological perspective, voice tone transformation in the video **"AI News Anchor Replaces an Al Jazeera Mubasher Presenter Live on Air and Interviews a Guest"** shows that the virtual presenter has an identity and looks more trustworthy, on screen.

The way she speaks is carefully planned to be calm well-balanced and clear. She uses controlled pacing, pauses and changes in pitch to sound like a news anchor. In a way these features show that she is neutral, dependable and skilled.

Accordingly, voice tone functions not merely as a medium for transmitting speech but as a symbolic resource that persuades viewers to attribute the qualities of a human news presenter to an AI-generated speaker.

The analysis also reveals the limitations of this semiological sign. Emotional variation remains restricted, and vocal rhythm is highly regular, reducing the sense of spontaneity and exposing the artificial nature of the performance, particularly during moments of dialogue requiring emotional responsiveness or immediate interaction.

Consequently, voice tone transformation aims not only to imitate human speech but also to construct a system of auditory signs that reinforces the virtual presenter's communicative authority within media discourse while retaining distinctive semiological differences that separate artificial performance from authentic human communication.

Semiotic Gap: The Semiotics of the "Uncanny Valley"

There is a difference between what we hear and what it really means when we listen to a machine talking. The machine sounds like a person. It does not really feel or talk like a person. This is what happens when the machine talks to **Dr. Taleb Al-Ashqar.**

The machine sounds like it is talking. It does not really understand how people feel. This is a problem because the machine can sound like a person. It is not really like a person.

This is what is happening with intelligence right now. Machines are getting better at talking like people. They are not really feeling like people.

This is what **Dr. Mohammed Al-Nashar** thinks too. He is the Dean of the Faculty of Language and Media at the Arab Academy, for Science, Technology and Maritime Transport.

"Despite scientific progress and the necessity of benefiting from it, the human dimension cannot be ignored in any profession, whether broadcasting, journalism, or medicine. Interaction between the presenter and the audience is extremely important."

He further explains:

"At the verbal level, every presenter has a unique style of conveying information through vocabulary choices and the way they engage rapidly with guests, particularly in talk shows. At the non-verbal level, facial expressions and hand gestures communicate happiness, sadness, and emotional engagement with events and news. Together, these elements determine the success and popularity of presenters and their programs."

According to Dr. Al-Nashar, this combination of verbal and non-verbal communication remains beyond the capabilities of current AI news presenters .

Ethical and Legal Dimensions of Voice Tone Transformation in Media

As artificial intelligence technologies become increasingly integrated into media institutions, growing concerns have emerged regarding the ethical and legal standards required to ensure their responsible use, particularly given their capacity to generate content that is often indistinguishable from human-produced media.

In this context, **Heba Helmy,** a lecturer at the Faculty of Mass Communication at **Al-Ahram Canadian University,** argues that integrating artificial intelligence into journalism requires the establishment of clear professional, ethical, and legal regulations to preserve credibility, accuracy, and transparency in media production.

She cites an incident that occurred in China at the end of March, during which two AI-generated virtual news presenters appeared on social media without any disclosure of their artificial nature. The presenters delivered

politically biased news content under the banner of a fictitious news channel named **Wolf News**.

According to Helmy, this incident illustrates the risks associated with exploiting AI technologies to manipulate public opinion or disseminate misleading political messages. imilar concerns have emerged as some Indian media organizations prepare to conduct an interview between a virtual news anchor and **Prime Minister Narendra Modi**.

Based on these developments, Helmy supports the growing calls by researchers and technology specialists for comprehensive regulatory and legislative frameworks capable of limiting the potential harms associated with artificial intelligence while ensuring its responsible application.

Within this context, the **European Union** continues to develop one of the world's first comprehensive legal frameworks governing artificial intelligence. The proposed legislation aims to establish clear obligations for AI developers while promoting the safe and ethical deployment of AI technologies across various sectors, particularly the media industry.

4. Results of the Analysis of the Media Models

The analysis of the three media models produced the following findings:

- The application of artificial intelligence in news presentation has reached an advanced level of technological maturity. The examined organizations successfully developed virtual presenters with high visual and auditory realism through the integration of digital facial generation, Text-to-Speech systems, and lip synchronization, resulting in strong audiovisual coherence.
- Across the three models, synthesized voices demonstrated clarity, stability, composure, rhythmic consistency, and accurate pronunciation. These characteristics reflect professional broadcasting standards and contribute to perceptions of neutrality, objectivity, and credibility. However, limitations remained in emotional expression, vocal intonation, and spontaneous interaction compared with human presenters.
- From a semiological perspective, voice tone has evolved beyond simple communication to become a sign of institutional authority, professionalism, neutrality, and credibility. The synthesized voice represents an institutional voice associated with the media organization rather than the individual presenter. This reflects a broader shift from human presence toward algorithmic presence, while limited emotional expression reveals the artificial nature of these systems.
- Comparative analysis showed differences according to media format. The Al Jazeera Mubasher election model maintained a formal and authoritative tone, while the Aisha Gaban documentary demonstrated greater tonal variation. The live interview model showed greater capacity to simulate television dialogue, although limited emotional responsiveness remained an important indicator of artificiality.
- Technically, realistic AI-based media presentation depends on integrating large language models, Text-to-Speech systems, facial animation, and lip synchronization. Despite significant progress, these technologies still struggle to reproduce human emotional nuance, spontaneous interaction, and contextual adaptation.
- Finally, the rapid development of AI-generated presenters raises important ethical and legal concerns. Transparency regarding AI use, protection against media manipulation, and clear professional and legal regulations are necessary to preserve credibility and ensure responsible use of AI in journalism.

5. Conclusion

This study concludes that artificial intelligence-based voice tone transformation technologies have evolved beyond merely enhancing vocal quality to become an essential component in the construction of media meaning and the production of significance within media discourse. The analyzed models demonstrate these technologies' ability to reproduce the professional characteristics of broadcast speech—including clarity, composure, neutrality, and consistency—thereby strengthening perceptions of credibility and professionalism.

The thing is, when we look at it closely we can see that there is a difference between what technology can do and how people really express themselves. This difference is clear when we listen to voices that are made by computers. They do not sound very natural because they cannot express a range of emotions or feelings in the way that people do. Sometimes they sound a bit fake. So when we hear a computer made voice it does two things at the time. It makes the message sound more important and trustworthy. At the same time it is clear that it is not a real person talking. This is what computer generated voices are, like. They have a kind of double meaning.

Accordingly, the future integration of voice tone transformation into journalism depends not only on continued technological innovation but also on the development of robust professional, ethical, and legal frameworks

capable of ensuring the responsible use of these technologies. Such frameworks are essential for maintaining public trust and safeguarding the credibility of media communication in an era of accelerating digital transformation.

6. References

Al-Ahram. (2023, May 3). AI presenter conquers the world: Where is the technology in media heading? Al-Ahram, 49821.

Alanaz, S. &. (2024). Exploring deepfake technology, creation, consequences, and countermeasures. Human-Intelligent Systems Integration.

Al-Qadri, M. A. (2024, October). Employing Deepfake and Voice Cloning Technologies in Creating Educational Videos Featuring Islamic Scholars. Journal of Sharia and Islamic Studies, 337.

Anis, I. (1999). Linguistic Sounds. Maktabat Nahdat Misr and its Press, 99.

Ashour, M. (s.d.). What is the relationship between tone and idea in persuasion? . Electronic Consultation Portal.

Barthes, R. (1967). Elements of semiology . Hill and Wang, pp. 9–17. (T. A. Lavers & C. Smith, Trad.)

Barthes, R. (1977). Image, music, text . Fontana Press. (T. S. Heath, Trad.)

Belaid, S. (s.d.). In Defense of the Language of the Media. High Council for the Arabic Language, 109. Algeria.

Biermann, J. (2025, February 4). Tone of voice in corporate communications: The right tone sets the brand. Empower.

Bouchair, K. (1980). General linguistics: Phonetics and phonology. 162. Egypt.

Bouchair, K. (2000). Phonetics. Dar Gharib for Printing, Publishing, and Distribution, 517. Cairo, Egypt.

Boukhaouch, S. (2019). The phonetic level in oral media performance. Journal of Phonetics, 110.

Channel. (2025, October). Channel 4 makes TV history with Britain's first AI presenter. Channel 4 Press.

Chen, H. X. (2025). Comprehensive exploration of diffusion models in image generation: A survey. Artificial Intelligence Review, Article 99, 58.

De Saussure, F. (1985). Course in general linguistics. Dar Afaq Arabia, 33. (T. Y. Ghazi & M. Al-Nasr, Trad.)

Faris, I. (1991). Standards of language. 1, 5. (Harun, Compilateur) Al-Jabal.

Gerard, A. (2023, August 23). From fiction to reality: A deep dive into voice cloning technology. Medium.

Goodfellow, I. B. (2016). Deep learning. MIT Press.

Iluwar, R. (1980). An introduction to linguistics. Damascus University Press, 86. (T. B. Al-Qasim, Trad.)

Kadam, S. J. (2024). ReVoice: A neural network based voice cloning system. . 2024 IEEE 9th International Conference for Convergence in Technology (I2CT).

Kim, J. K. (2021). Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech. Proceedings of the 38th International Conference on Machine Learning, 5530–5540.

McLuhan, M. (1964). Understanding media: The extensions of man. McGraw-Hill.

Mubasher., A. J. (2024). AI presenter on Al Jazeera Mubasher launches special coverage of the 2024 U.S. presidential election. YouTube.

Mubasher., A. J. (s.d.). AI presenter replaces. Al Jazeera Mubasher presenter live on air and interviews a guest. YouTube .

Mujahid, A. K. (1985). Linguistic meaning in the Arab tradition. Dar Al-Diyaa for Publishing and Distribution, 166. Amman, Jordan.

News., S. (s.d.). An AI-generated woman presents a documentary on British television. SNRT News.

Okasha, M. (2005). Linguistic Analysis in Light of Semantics: A Study of Phonological, Morphological, Syntactic, and Lexical Semantics. 1, 17–18. Cairo, Egypt: lil-Jami'at.

Omar, A. M. (1993). Contemporary Arabic language errors among writers and broadcasters. Alam Al-Kutub, 2nd, 39. Cairo, Egypt.

Peirce, C. S. (1931–1958). Collected papers of Charles Sanders Peirce . Harvard University Press, 1–8, 228–229.

Radford, A. K. (2022). Robust speech recognition via large-scale weak supervision. OpenAI.

Saussure, F. d. (1983). Course in general linguistics. Duckworth. (T. R. Harris, Trad.)

Silverman, C. (2020). Verification handbook for disinformation and media manipulation . Al Jazeera Media Institute, 180. (T. M. Zidan, Trad.) Qatar.

Soundeals. (s.d.). The art of portraying emotions through sound: Can sound alone tell a complete story?

Tamam, H. (1980). Methods of research in language. 160. Egypt.

Van den Oord, A. D. (2016). WaveNet: A generative model for raw audio. Proceedings of the 9th ISCA Workshop on Speech Synthesis Workshop (SSW 9), 125.

Wang, Y. S.-R. (2017). Tacotron: Towards end-to-end speech synthesis. Interspeech, 4006.

Widdowson, H. G. (1996). Linguistics. Oxford University Press, 53–61.

Willayid, A. (2003). Applied linguistics in teaching Arabic to non-native speakers. Al-Jawhara Publishing and Distribution, 1st, 128. Jordan.

Youm., E. A. (s.d.). Virtual journalist Aisha Gaban. Facebook video.

Zain, M. A. (2005). Linguistic analysis in the light of semantics: A study of phonological, morphological, syntactic, and lexical semantics. Dar Al-Nashr Lil-Jami'at, 17–18. Cairo, Egypt.