

THE PROBLEM OF NATURALITY OF INTONATION IN AI SPEECH SYSTEMS

Cho'lliyeva Gulchehra Toshpo'lot qizi

Abstract: This article explores the issue of naturalness in intonation within AI-based speech systems. While artificial intelligence can generate grammatically accurate speech, it still struggles to reproduce the emotional and pragmatic depth inherent to human communication. The paper discusses the role of intonation as a cognitive and communicative marker, emphasizing its function in conveying meaning, rhythm, and emotion. It analyzes the limitations of current prosodic models, cultural and linguistic variability, and ethical concerns surrounding voice cloning and identity. Furthermore, the article outlines future directions for enhancing naturalness in AI-generated speech, including contextual prosodic planning, emotional modeling, and multimodal integration.

Keywords: intonation, prosody, artificial intelligence, speech synthesis, naturalness, emotion, pragmatics, acoustic modeling, communication, ethics.

Аннотация: В статье рассматривается проблема естественности интонации в системах речи, основанных на искусственном интеллекте. Несмотря на грамматическую правильность машинной речи, ей по-прежнему не хватает эмоциональной выразительности и человеческого звучания. Анализируется значение интонации как когнитивного и коммуникативного индикатора, обеспечивающего передачу смысла, ритма и эмоций. Особое внимание уделено трудностям моделирования просодии, культурным и языковым различиям, а также этическим вопросам, связанным с клонированием голоса и идентификацией личности. В заключение обозначены перспективные направления — контекстное планирование просодии, эмоциональное моделирование и мультимодальная интеграция.

Ключевые слова: интонация, просодия, искусственный интеллект, синтез речи, естественность, эмоции, прагматика, акустическая модель, коммуникация, этика.

Today we vote with controllable technologies during We live in . Say "Hey Siri" to your phone . when we do or from Google Assistant the weather When we ask , artificial intellect " Hear us , " " Answer us. " gives ” and as if human-like communication does . But this in communication the most thin , but the most important thing — tone , that is intonation — still perfect at the level adapted no . Grammatically right , but emotional " cold " or " artificial " sounding speech technology the most big from the shortcomings is one . Tone only sound vibration not , it is the idea their heart , their feelings voice , humanity symbol .

Human in his speech every one word , every one pause to the point service "I am here . " said simple of the word himself/ herself various in tone if you say so , completely other content profession will : doubt , pride , joy or regret . Human in the brain this melody automatic is formed ; it is an idea and of feeling natural Artificial intellect for and this process – complicated algorithmic problem



Date: 17th October-2025

AI systems the text to the sound turn takes , but meaning , context and feelings feel the harmony can't . That's why for " grammatical " " righteousness " and " humanity " between "naturalness " deep difference there is .

Speech synthesis systems , mainly tw in stages works : first , the text linguistic signs and phonetic to the information divides , then him/her voice to the shape In this process prosody , that is melody , rhythm , accent and pause such as elements to the text soul But every one in the language , every one in culture melody system different : English in the language interrogation melody with Uzbek in the language interrogation melody or Chinese from the tone consists of accent system to each other It doesn't look like that . Therefore, on a global scale "universal tone" for working AI "model" create almost It is impossible .

Ohnag's natural not to come out one how many reason there is : first , the body information often emotional in terms of neutral from the records consists of As a result , the model produces " emotionless " speech . learns ;

secondly , the meaning and melody between connection algorithmic in a way complicated . In the sentence intention is " emotional " context , " it to the number convert easy not ;

thirdly , the human variable mood and cultural in terms of formed speech habits car for in advance prophecy as impossible are elements .

Today artificial intellect speech melody naturalness increase for different the roads try is seeing . Some systems neuron networks feel them through analysis does , others F0 (pitch) and rhythmic the lines human from the speech learns . Growing up bodies , various in languages recorded audio data and multimodal analysis (text , sound , image) together to study) this on the road big step It was also multichannel . study methods (e.g. , previous words context analysis doer models) intonation logical in consistency storage opportunity is giving . But the problem is still complete solution done not . Because algorithms does not "feel " emotions , they only mathematician patterns repeats . Therefore for " humanity " in the tone create For , AI is not only acoustic models , maybe pragmatic understand the meaning need .

AI voice natural from the exit besides , moral There are also issues . Current on the day human his voice cloning through fake individuals create it is possible , this and identification the risk brought is issuing . From this besides , different in cultures of the tone one kind not happened interpretation problems further deepens : one in the language sincere melody in another rudeness as acceptance to be done possible . Therefore AI speech for in systems intonation study not only technical , maybe ethics , sociology and cultural studies also related to is a matter .

Near AI technologies in the years development with intonation naturalness increase on the way new directions appearance to be expected :

- a) to the context suitable prosodic planning — artificial of the system text meaning and conversation the tone together analysis to do ;
- b) emotional modeling — artificial the speech human to the senses zoom in ;



Date: 17th October-2025

c) multimodal integration — voice , face expression , gestures and body language with together analysis .

Intonation naturalness improvement this only technician achievement This is not a person . and car between communication further sincere , trustworthy and natural thrown into action is a step .

AI speech systems development human speech deep analysis without doing imagination as It won't be . Ohnag this only of sound change not , maybe thought , feeling and personality acoustic It is an expression of this . for artificial in the process of " humanizing " the intellect the tone study technology , linguistics and philosophy intersected in place standing the most current is the direction .

REFERENCES USED:

1. Chafe, W. (2018). The Relationship between Thought and Language: Intonation as a Cognitive Structure. New York: Routledge.
2. Levelt, W. J. M. (1999). Models of Speech Production. Cambridge University Press.
3. Makarov, D. & Belousov, A. (2021). Artificial Intelligence in Speech Synthesis: Challenges and Prosody Modeling. Moscow: RUDN Press.
4. McKeown, G. (2023). Emotion in Speech Technology: Towards Humanized AI Voices. London: Springer.
5. Zhang, X., & Li, H. (2022). Prosodic Modelling for Natural Speech Synthesis in Multilingual Contexts. IEEE Transactions on Audio, Speech, and Language Processing.
6. Patel, A. D. (2010). Music, Language, and the Brain. Oxford: Oxford University Press.
7. Taylor, P. (2009). Text-to-Speech Synthesis. Cambridge: Cambridge University Press.
8. Юлдашева Д.Н., Юсупова А.Ш., Чўллияева Г.Т. О невербальных средствах речи. – BuxDU Ilmiy axboroti, 2023. –4-son. – 108-115-b.