Smart Media Journal / Vol.15, No.2 / ISSN:2287-1322
https://dx.doi.org/10.30693/SMJ.2026.15.2.56

비식별 캔들코드 기반 음성변이의 디지털 감정 시각화

(Digital Emotion Visualization of Voice Variability Using De-identified Candle Code Representation)

강민구*
(Mingoo Kang)

요 약

본 논문에서는 음성 기반 감정 분석의 개인정보 노출 문제와 비언어적 감정신호 손실 문제를 동시에 해결하기 위해, 비식별 음향특징(ΔHz, ΔdB, Δms)을 활용한 이진 감정코드(Binary Emotion Code, BEC)의 생성방식과 이를 캔들코드(Candle Code) 형태로 시각화하는 새로운 감정변이 분석 프레임워크를 제시한다. 이는 기존 STT(Speech-To-Text) 기반 감정분석은 언어 의존성과 재식별 위험이 크며, 감정의 시간적 흐름(증가·감소·유지)을 정량적으로 표현하도록 제안하였다.

이로써, (1)HNR, Pitch, Intensity, Jitter, Shimmer 등 7종 음향특징을 100[ms]단위로 추출하고 정규화한 뒤, (2)임계값 기반 3비트 이진코드로 감정 변동을 BEC로 부호화하고, (3)이를 캔들형 시각화(Open - Close - High - Low)로 표현하였다. 웹 기반 시험에서 감정 상승·하강 패턴을 스펙트로그램 대비 더 빠르게 인지할 수 있음을 확인하였다. 제시된 방식은 발화자의 음색·포먼트 정보를 포함하지 않아 비복원성이 보장되며, 프라이버시 친화형 실시간 감정 모니터링에 활용 가능하다.

■ 중심어 : 디지털 감정분류 ; 캔들코드 ; 음성변이 ; 비식별 감정모니터링 ; 이진 감정코드(Binary Emotion Code, BEC)

Abstract

In this study, a new emotion_variation analysis framework is presented to simultaneously address the issues of personal information exposure and the loss of paralinguistic emotional signals in speech_based emotion recognition. To achieve this, de-identified acoustic variation features—ΔHz(frequency), ΔdB(intensity), and Δms(duration)—are employed to generate a Binary Emotion Code(BEC), and the resulting emotional transitions are visualized through a Candle Code representation. This approach overcomes the linguistic dependency and high re-identification risk inherent in conventional STT(Speech-to-Text)_based emotion analysis and enables the temporal flow of emotional changes(increase, decrease, and neutrality) to be quantified.

Accordingly, (1)seven acoustic features including HNR, Pitch, Intensity, Jitter, and Shimmer are extracted and normalized at 100[ms] intervals, (2)emotional variations are encoded into a threshold-based 3_bit BEC, and (3)the encoded changes are visualized using candle_style components(Open - Close - High - Low). Through web_based evaluations, emotional rise-and-fall patterns are observed to be identified more rapidly compared to spectrogram_based visualization. Because the method excludes identifiable information related to timbre and formant characteristics, non-invertibility is ensured, enabling practical application to privacy_preserving, realtime emotion monitoring.

■ keywords : Digital Emotion Classification ; Candle Code Visualization ; Voice Variability (Speech Variability) ; De-identified Emotion Monitoring ; Binary Emotion Code (BEC)

Ⅰ. 서 론

최근 스마트 센서 및 인공지능 기반 음성처리 기술의 발전으로, 음성신호를 이용한 각성도(Arousal) 모니터링 연구가 활발히 이루어지고 있다[1,2].

이러한 음성 기반 감정인식은 스마트 디바이스와 인공지능과 함께 다양한 분야에서 활용되고 있다.

특히 각성도(Arousal)는 건강 모니터링 등에서 중요한 감정 지표로 활용된다. 그러나 기존 연구는 다음 두 가지 근본적 한계를 갖는다.

(1) 언어 의존성과 비언어 신호 손실

STT 기반 감정분석은 텍스트 의미분석인 BERT(Bidirectional Encoder Representations from

* 정회원, 한신대학교 AI.SW대학
이 논문은 2025년 한신대학교 학술연구비 지원에 의하여 연구되었음.

접수일자 : 2025년 11월 06일
수정일자 : 2025년 12월 01일
게재확정일 : 2025년 12월 12일
교신저자 : 강민구 e-mail: kangmg@hs.ac.kr

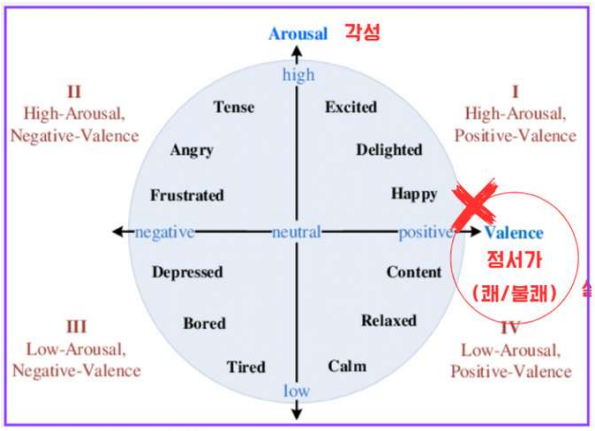
Transformers)에 강점을 갖지만, 억양·속도·리듬·발성강도 등 비언어적(Paralinguistic) 감정 신호를 반영하지 못한다.

(2) 재식별(Re-identification) 위험

스펙트로그램·파형에는 화자의 음색(Formant), 주파수 패턴, 성대특성 등이 포함되어 있어 개인정보가 노출될 수 있다.

따라서 언어 독립적·비식별 기반·시간적 감정변이를 시각화하는 새로운 프레임워크의 필요하게 된다.

본 연구는 위의 문제를 해결하기 위해 STT 제거로 언어 의존성 해소, 비식별 음향특징(ΔHz·ΔdB·Δms) 추출과 3비트 감정코드, 캔들코드 기반 감정 시각화를 포함한 프라이버시 친화형 감정변이 분석 구조를 제시한다. [그림1]은 감정변이 분석을 위한 Russel각성도 모델의 1차원 음성변이분석을 위한 차원축소이다[1].



Russell(1980)감정차원 모델(Circumplex Model)
• Valence: 쾌감/불쾌감 정도(무시>1차원 축소)
• Arousal: 낮은 각성에서 높은 각성

(Principles of Neural Science (6th ed.), Kandel et al., McGraw-Hill, 2021)

그림1. Russel의 각성도 모델을 기반으로 한 1차원 음성변이 시간적 분석모델 제시

[그림1]처럼 시간적인 감정변화에 대한 1차원적 각성도 분석 만을 위해, 발화음성의 주파수·음압·시간적 변동을 실시간으로 분석하고자 한다[1-6].

이때, 비언어적 음향특징 3요소를 정규화를 제시한다. [표 1]은 각 음향요소의 변화량을 BEC(Binary Emotion Code)로 표시하고자, '말 떨림/발성시간'의 실시간적 분석을 위해 음성변화량을 계량화한다[7].

표1. 음향요소 분석을 통한 ΔHz·ΔdB·Δms 정규화 기반 음성 데이터 변화량 산출분석[7]

음향요소	음성 데이터1			음성 데이터2		
	최고값	최저값	변화량	최고값	최저값	변화량
음량(dB)	a11	a12	ΔdB1	a21	a22	ΔdB2
주파수(Hz)	b11	b12	ΔHz1	b21	b22	ΔHz2
발성시간(ms)	Δms1			Δms2		
말떨림	(ΔdB&ΔHz)1			(ΔdB&ΔHz)2		

이때, 각 음향적 특징요소를 사전 정의된 임계값(Threshold)과 가중치(Weight)에 따라 이진 감정코드로 변환하고, 이를 [그림2]처럼 캔들코드(Candle Code) 로 시각화하였다[5].

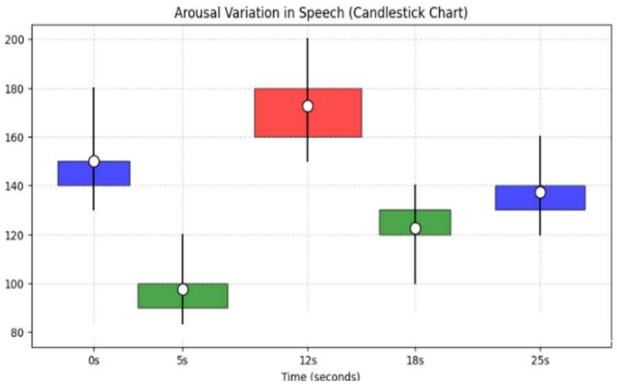


그림2. STT 없는 음성변이 기반 캔들차트의 캔들코드 시각화

- 녹색캔들: 음성변이 기준선(Baseline,Neutral상태)
- 파랑캔들: 기준선 대비 음성 변이가 감소(0)
- 적색캔들: 기준선 대비 음성 변이가 증가(1)

이러한 캔들코드 시각화는 음성신호 내 감정변화 시작(Open), 종료(Close), 최대(High), 최소(Low)의 4개 요소로 재구성하여 시계열 패턴을 표시한다.

이로써, 비식별화된 발화자 음성 데이터도 감정 흐름의 정량적 변화를 직관적으로 표현할 수 있다.

이와 같이, ΔHz·ΔdB·Δms 변화 값을 기반으로 한 캔들코드 시각화는 [그림2]와 같이 임계 값 기반 캔들차트(Chart)처럼 색상 매핑(Color Mapping)을 통해 감정 변화의 시간적 방향성과 강도를 동시에 표현할 수 있는 특징이 있다.

시험용 사이트는 발화자가 비식별화된 감정변이를 직관적인 캔들코드가 비복원적(Non-invertible)이며, 비식별 상태에서도 정량적 각성도 추정이 가능하지만 발화자 복원이 불가능할 수 있다.

Ⅱ. 음성기반 감정분류와 캔들코드 할당

최근 음성 감정인식 연구는 음성 신호의 스펙트럼, 주파수, 강도(Intensity)등을 입력으로 하는 CNN, LSTM 기반의 딥러닝 분류 모델을 중심으로 발전해 왔다[1-6].

그러나 이러한 접근은 모델의 복잡도와 재식별 위험(Re-identification Risk), 그리고 시간적 감정 흐름(Time-series Emotion Dynamics)에 따른 감정변화의 분석력 부족이라는 한계를 지닌다.

이에 따라 비식별 음성인식 기반 감정분석(De-identified Voice Emotion Analysis) 연구로 EASY, DEEMO 등의 모델은 MFCC·Spectral Flux 등 저차원 음향적 특징만을 남겨 발화자 신원을 제거하면서 감정인식을 수행하고 있다. 이러한, EASY, DEEMO는 MFCC, 스펙트럴 플럭스 등은 저차원 특징만을 남겨 음색·포먼트를 제거하는 비식별 기법을 제시하고 있다[3-4].

하지만, 이들 연구는 STT기반 감정인식의 한계, 텍스트 중심 분석으로 비언어 신호반영 부족, 언어 종속성과 STT오류에 취약 음성 - 텍스트 변환 과정에서 신원특징의 보존하는 이슈가 존재한다.

본 연구에서는 이러한 한계를 보완하기 위해, 비식별화된 음향특징(HNR, Pitch, Intensity, Jitter, Shimmer) 을 정규화한다.

이후에 그 변화를 3비트로 변환한 뒤, 시간 축의 감정 변동을 캔들코드(BEC, Binary Emotion Code) 형태로 시각화하는 방식을 제시하고자 한다. 이러한 캔들코드는 주식 거래장에서의 캔들차트(Candle chart)원리를 응용하여 감정의 상승(↑), 하강(↓), 유지(=)를 색상, 길이, 기울기로 표현하도록 하였다. 이로써, 기존의 STT 및 BERT 감정모델의 비의존적이면서도 프라이버시 친화적인 비식별적 감정변이를 시각화하고자 한다.

2.1. STT 기반 감정인식 연구 동향 분석

최근 음성 감정인식 연구의 다수는 STT(Speech-To-Text) 기술을 이용하여 발화 내용의 언어적 의미(Emotion Feature)를 감정을 추정한다.

BERT 모델처럼 텍스트 기반 감정분석 모델(EASY, EmoSpeech 등)은 자연어 감정어휘 사전과 음성-텍스트 매핑 알고리즘을 결합하여 감정분류 정확도를 향상시켰다.

이러한 방식들의 장점은 언어적 정서 표현을 정밀하게 포착할 수 있다. 또한, 대화형 인공지능·콜센터 감정분석 등에서 실효성이 검증되었다. 그러나 이러한 STT와 텍스트 기반 감정분석 접근은 다음과 같은 한계를 가진다.

첫째, 발화자의 음색, 발성강도, 주파수 변동 등 비언어적 감정요소(Paralinguistic cues)를 반영하기 어려웠다.

둘째, STT 인식오류(Word Error Rate, WER) 와 언어 종속성(Language Dependence)으로 인해 실제 감정추론의 일관성이 저하되었다.

셋째, 텍스트화 과정에서 발화자의 신원정보·억양패턴 등 개인정보 노출 위험이 존재한다[6-10].

본 연구에서는 STT를 완전히 제거하고, 순수 음향특징 기반의 비식별 감정분류 구조로써 언어 의존성과 재식별 가능성을 동시에 차단하였다.

2.2 비식별 음성분석과 발화자 익명화 분석

비식별 음성분석(Voice De-identification) 연구는 개인정보보호법 강화에 따라 급속히 발전하였다.

EASY 프레임워크(Emotion Analysis without Speaker Identity), DEEMO(De-identified Emotion Modeling)는 MFCC(Mel-Frequency Cepstral Coefficients), 스펙트럼 변화율(Spectral Flux), 음성 품질 지수(Voice Quality Index) 등의 통계적 특징만으로 발화자 신원을 제거하면서 감정 정보를 표현하였다[3][4].

이들은 가우시안 혼합모델(Gaussian Mixture

Model, GMM) 또는 변이자동부호화(Variational Autoencoder, VAE) 기반으로 발화자 음색분포를 평균화(Averaging)하여 익명화하였다.

하지만, 이러한 비식별 모델은 복잡한 후처리와 모델 파라미터 조정이 필요해 실시간성(Real-time Capability)이 낮고, 시각적 해석(Interpretability)이 부족하다.

[표2]는 복잡한 음향모델을 사용하지 않고, HNR·Pitch·Intensity·Jitter·Shimmer 등의 단순 물리량을 8비트(대분류>감정정보 3비트 + 중분류>감정정보 3비트 + 음향요소>2비트)로 정규화한다. 이러한 디지털 감정 부호화(Binary Emotion Coding)를 통해 감정분석의 실시간 시각화가 가능하게 되었다[7].

표2. 음향요소와 감정정보 모니터링 감정부호 8비트 할당설계

대분류 감성정보 (코드할당 : 3비트)		중분류 감성정보 (코드할당 : 3비트)	음향요소 (2비트)
Pleasure	000	glad (000), happy (001), pleased (010)	음량 변화량 (ΔdB)
Excitement	001	excited (000), delighted (001)	
Arousal	010	astonished (000), aroused (001), alarmed (010), tense (011), angry (100)	주파수 변화량 (ΔHz)
Distress	011	afraid (000), annoyed (001), distressed (010), frustrated (011)	
Misery	100	miserable (000)	발성시간 (Δms)
Depression	101	sad (000), gloomy (001), depressed (010)	
Sleepiness	110	bored (000), droopy (001), tired (010), sleepy (011)	말떨림 (ΔdB&ΔHz)
Contentment	111	calm (000), relaxed (001), satisfied (010), at ease (011), content (100), serene (101)	

2.3 시계열 감정 시각화 모델 분석 : 파형·스펙트럼·운율곡선과 캔들코드

감정 시각화는 주로 파형(Waveform), 스펙트로그램(Spectrogram), 운율곡선(Prosody Curve) 등으로 감정 변화를 표현해 왔다.

[표3]에서는 기존 시각화 대비, '이진부호(3비트/8비트 코드할당)+캔들형 시각화' 구조를 결합한 감정변이의 시각적 추세를 시각화하고자 하였다.

이러한 방식들은 시간·주파수 변화를 정밀하게 보여주지만, 감정의 방향성(증가/감소)과 세기변화(Amplitude Trend)를 직관적으로 인식하기 어렵고, 복잡한 스펙트럼 형태를 갖는다[5-6].

표3. 음성변이 요소의 모니터링을 위한 프로세스 설계/분석

모니터링	음성요소 분석과 설계의 특징분석
음성입력	스마트폰을 통해 약 10초 길이의 짧은 음성 구간 수집
특징추출	실시간으로 6가지 음향 특징 추출 (HNR,강도,지터,피치,쉬머, 발화속도)
정규화	중앙값/사분위범위 Z-점수 정규화
각성도점수	평균 점수 계산 (척도: 1~5)
시각화	캔들 코드 및 선형 차트 시각화

이에 비해 [표4]는 제시한 캔들코드는 각 음성 프레임의 ΔHz, ΔdB, Δms 변화를 네가지 요소로 매핑하여, [그림2]처럼 감정 변동의 강약과 방향을 3색(녹·청·적)으로 시각화한다.

표4. 감정상태와 캔들코드의 이종 시간화 차별성 표시 비교분석

감정 상태	캔들 표시	감정상태 변이
Excitement	▲ 상승형(↑)	에너지 급상승
Calm	━ 안정형(=)	평온한 상태
Anxiety	▼ 하강형(↓)	불안·긴장 증가
Arousal	⬆ 부분 상승	경미한 각성

이로써 [표5]처럼 기존 EASY, DEEMO, 스펙트럼 분석 연구의 한계를 극복하고, 감정의 시간적 연속성과 직관적 표현할 수 있게 되었다.

표5. 음성인식 기반의 감정인식 및 캔들코드 시각화 비교 분석

비교항	특징	장점	한계점	Candle Code
STT 감정인식	언어적 맥락기반 감정추론	의미 단위 분석 용이	비언어 정보미포함 언어 의존성	STT 제거 언어독립
비식별 음성분석 (EASY, DEEMO)	신원정보제거후 감정유지	개인 정보보호 강화	복잡한 모델·저속처리	단순 물리량 이진화,실시간 처리
파형/스펙트럼 시각화	주파수·세기 시각화	세밀한 시간·주파수 해상도	직관성 낮고, 발화자 난해	3비트 캔들코드 감정흐름 색상·패턴

[표6]은 비식별 감정분석 시스템이 갖는 재식별(Re-identification) 위험노출을 최소화 하도록 캔들코드로 개인정보의 비복원성을 강화하였다.

표6. 비식별 감정분석 환경의 재식별 위협과 대응 비교분석

위협요소	특징분석	대응방식 분석
음색 복원 공격 (Timbre Reconstruction Attack)	음성의 포만트(Formant) 특성을 역추적하여 발화자 재식별	음향특징 집합 제한(HNR, Pitch, Jitter 등 1차 파라미터만 사용)
모델 반추출 공격 (Model Inversion)	감정모델 파라미터로부터 개인 데이터 복원	감정코드 이진화 및 차등값 분산처리
데이터 연계 공격 (Data Linkage)	다중 사이트의 음성데이터 연계 익명성 상실	세션단위 처리 후 로그 삭제, 서버 저장 최소화
시각화 역추적 (Visualization Inversion)	시각화된 감정 패턴을 통한 재식별	캔들코드비복원성 (Non-reversible Mapping) 보장

[표5/6]에서 제시된 감정부호의 캔들코드의 시각화로 디지털인 압축하고, 시각화한 결과를 비복원형(Non-invertible)으로 유지할 수 있다. 이로써, 감정분석 정확도와 프라이버시 보호를 달성하였다.

III. 웹 기반 음성변이 추정과 캔들코드

본 절에서는 발화 음성의 '말 떨림(Jitter)', '발성시간(Duration)'에 대한 음성변이 추정으로 사이트를 통해 검증하였다. 이때, 입력신호는 10초 단위로 수집되며, 100 [ms]프레임 단위로 분석하게 된다.

표7. ΔHz,ΔdB,Δms기반 비식별 감정변이와 3비트 코드할당

음향요소 분류특징	감정변이 특징매핑	코드할당
ΔHz (Pitch Variation)	긴장/흥분 변화량	1st bit (0*/1**)
ΔdB (Intensity Variation)	강도와 에너지변화	2nd bit (0*/1**)
Δms (Speech Duration Variation)	발화속도·리듬변화	3rd bit (0*/1**)

* 저 각성도의 비트할당사례, ** 고 각성도의 비트할당사례

3.1 음성변이 부호화와 각성도 상태 산출식

[표7]은 실시간 음성변이 추정을 위해 제시된 3비트 캔들코드(Candle Code) 구조를 보여준다.

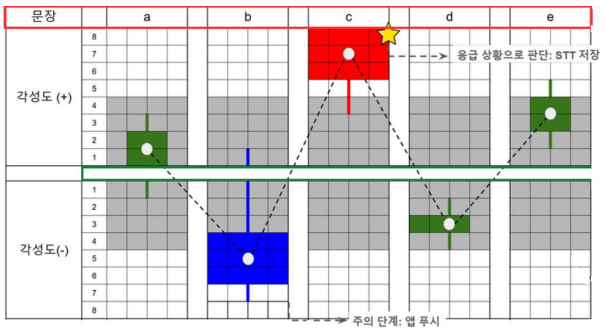


그림3. 각성도 변화문장에서 8단계(3비트)음성변이의 캔들코드

[그림3]은 각성도의 변화에 따른 문장에서 8단계의 음성변이를 3비트 캔들코드로 표현하고자 중간/중립, 하부/강하, 상부/상승 사례를 시각화하였다.

각성도의 시간적 변이를 표현한 부호는 시간 t에 따른 음성특징 xi(t)를 기반으로 상황인지 가중치 wi(t)와 보정항ε을 포함하는 동적 모델로 표현된다.

$$x_i'(t) = w_i(t) \times x_i(t) + \epsilon$$

$$\Delta\text{Feature}(t) = [\Delta Hz, \Delta dB, \Delta ms]$$

Binary Emotion Code=[code ΔHz, code ΔdB, code Δms]

$$\text{robust\_z}_i = \frac{x_i - \text{median}_i}{IQR_i}$$

$$\text{Arousal Score} = \text{round}(\text{mean}_{robustz} \times 1.0 + 3)$$

$$\text{score} = \max(1, \min(5, \text{score}))$$

- $x_i(t)$: 시간 t 에서의 음향특성 값
- $\text{median}(x_i)$: 중앙값
- $IQR = Q3 - Q1$: 사분위수 범위

윗 식에서 발화자 상황에 따라 음성특징(Feature) 변수 xi(t)는 상황인지 가중치 wi(t) 곱이다.

이러한 음성특징은 중앙값과 사분위수(IQR)로 강건 Z-점수(Robust Z-Score)의 정규화를 진행한다. 이때, 정규화된 각성도값은 평균을 기준으로 통합되어 각성도지수(Arousal Score,1~ 5)로 변환된다[4].

3.3 비식별 음성 각성도의 캔들코드 시각화

[그림 3]은 변이 변화의 음성변이 분류 및 코드할당 사례이다. 이로써 음성 감정 캔들코드는 영상·텍스트와 결합한 멀티 모달로 확장할 수 있다.

그림4 각성도 8등급 대·중·소 증상 분류와 3비트 코드할당
(저 각성도 이상:0~2, 정상:3~4, 고 각성도 이상:5~7)

[그림4]는시간 음성변이 기반 각성도 8등급의 대·중·소 증상 분류와 3비트 시각화한 사례이다.

$$C(t) = f(Binary\ Emotion\ Code(t))$$

각성도감소(↓): 0~2,안정(=):3~4, 각성도상승(↑):5~7

3.4 웹 기반 음성추출 시험분석과 고찰분석

[그림 4]는 시각화 분석사례로 웹 기반 시험 사이트(https://gaksungdo.vercel.app/test)의 화면이다.

이때, 10초단위로 발화입력을 100[ms] 프레임으로 처리한 결과이다. 이로써 캔들코드가 기존 파형 대비 감정상승·하강 변동을 직관적인 구분할 수 있다.

[그림4]처럼 시험 사이트의 사례분석으로 개인정보 노출 없이 음성분석 자료를 표시하였다[11 - 13].

그림에서는 피치(Pitch) 평균·변동성, 음성강도(Intensity)의 평균·동적범위, 발화 속도(Speech Rate), 강건 Z-점수 결과를 실시간으로 표시된다.

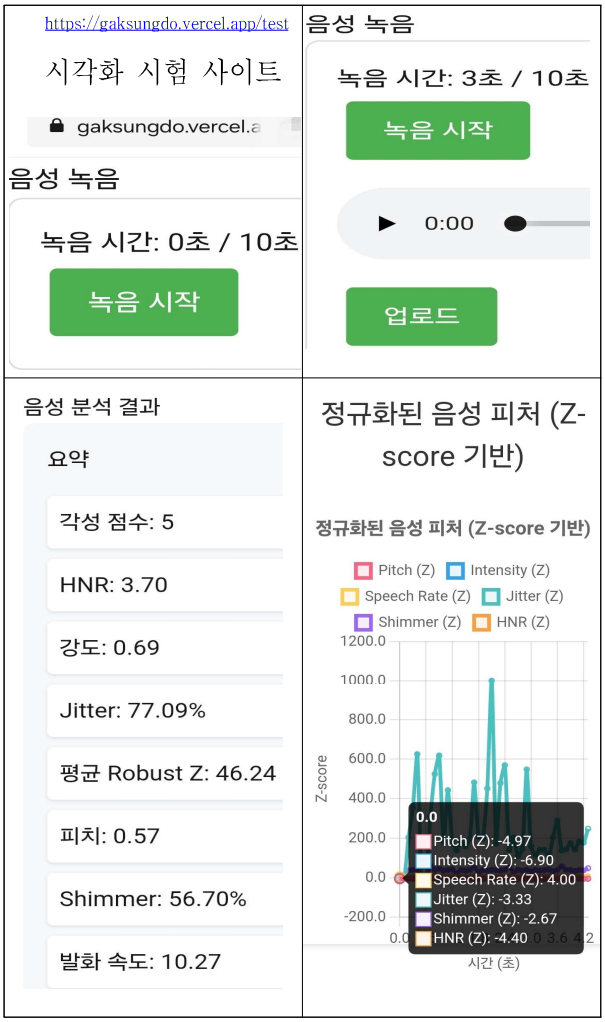


그림5 음성변이 평가용 시험 사이트에서 특징분석 예시화면

[그림5는 음성변이 평가용 시험 사이트의 분석한 예시 화면을 캡쳐한 것이다. 사이트에서 표출되는 음성특징 변수들은 아래와 같다.

- 고조파대잡음비(HNR):주기/비주기 성분비율
- 음성강도(Intensity): 음성 에너지와 음량
- 성대진동주기 불규칙성(Jitter): 주파수안정
- 강건 Z-점수(Robust Z-score) 표준화 : 음성 특성 값들의 적용으로 데이터의 신뢰성 향상
- 음의높이(Pitch):기본 주파수 추출과 음높이
- 발화속도(Speech Rate):음성구간/발화 빈도
- 음성강도 불규칙성(Shimmer):진폭변동/안정

[그림6]은 비식별 음성입력으로 감정 상태의 시각화하기 위한 코드를 정의하고 있다.

```
Input: audio 10s, frame=100ms
for each frame t:
  x = extract_features(audio[t])  # HNR, Pitch, Intensity, Jitter, Shimmer, SR
  z = robust_z(x)                 # (x - median)/IQR
  s_t = clamp(round(mean(z) + 3), 1, 5)   # Arousal Score
  b_t = binarize(z, thresholds)           # 3-bit code
  draw_candle(b_t, s_t)
```

그림6. 비식별 음성의 감정상태로 특징변수와 시각화 위한 코드

IV. 결 론

본 연구에서는 (1) STT 제거로 언어 의존성 해소, (2) 비식별 음향기반 감정코드 모델 제시, (3) 캔들코드 시각화를 통한 3가지 문제(식별 위험, 설명력 부족, 시각화 부재)를 구조화하였다.

이러한 암호화·비식별화를 기반으로 개인정보를 보호하는 프라이버시 친화형 AI 감정분석 프레임워크는 다음과 같이 정의하고자 하였다.

첫째, STT(Speech-To-Text) 비의존 음성 감정분석 구조를 제시하였다. 발화 신호의 HNR, Pitch, Intensity, Jitter, Shimmer 등 핵심 음향특징을 실시간으로 추출한다. 이를 강건 Z-점수의 정규화로 감정코드로 변환하는 과정은 아래와 같다.

이때, 입력 음성은 MFCC 변환 이전 단계에서 세그멘테이션과 비식별 적용되어 화자 음색·발화자 식별정보가 제거되었다. [표8]은 음성기반의 특징추출과 각성도 지표를 비식별 영역에서 수행되었다.

- **캔들코드 시각화 절차(Process)**

  음성입력 → 특징추출(Feature Extraction)→ 정규화(Normalization)→ 각성도평가(Arousal Scoring) → 캔들코드 시각화 → 비식별화(De-identification)

둘째, 제시된 캔들코드(Candle Code) 시각화는 감정의 상승(↑)·하강(↓)·유지(=) 패턴을 색상·길이·기울기로 표현한다. 이로써 시간적 감정 변동의 직관적 이해를 가능하게 되었다.

셋째, 웹 기반 시험 사이트 결과는 기존의 EASY, DEEMO 등 모델과 달리 스피커 임베딩 제거 과정 없이도 비식별화로 개인정보 유출 위험을 차단함으로서 프라이버시 보호를 제시하였다.

넷째, 캔들코드 기반 비식별 감정분석 프레임워크는 기존 스펙트로그램이나 선형 그래프 대비 감정의 시간적 변이와 각성도 변화(3비트/8비트 코드할당)를 직관적인 시각화표현이 가능할 수 있다.

표8. 음성인식 기반의 비식별 감정분석 비교결과 분석

구분	기존 방식 (시간화·시각화)	장점	한계점	캔들코드	장점	차별성
STT 기반 감정 분석	발화 음성을 텍스트로 변환 후 감정 단서 추출	언어적 의미 분석 용이	비언어적신호 (억양·속도) 손실, 개인정보 노출 위험	비식별 이진 부호화	개인정보 보호, 비언어적 특징 반영	STT 불필요, 프라이버시 보장
스펙트럼	시간-주파수 분포를 색상 표현	전문가용 분석에 정밀	복잡, 일반인 직관성 낮음	색·막대	직관적, 비전문가 이해 가능	시각적 단순화 미디어 친화적
파형분석 (Waveform)	시간축에 진폭 변화를 표시	기본 음성 구조 파악 용이	감정·맥락 표현 부족	증가/감소	변동성, 각성도 직관적 표현	파형보다 감정 변화 해석 강화
운율곡선 (Pitch/Intensity Curve)	피치·강도의 시간적 변화 곡선	억양·리듬 분석에 강점	데이터 해석에 전문지식 필요	다중요소 집약	HNR·Pitch·Jitter ·Shimmer 등 복합	다변량 통합표현
정량 점수화 (Arousal Index)	평균값 기반 단일 점수화	정량 비교 용이	변화 추세 표현 부족	동적 시각화	시간축 따라 증감 시각화	정량+정성 동시표현

이러한 직관적이고 비식별화된 캔들코드 시각화는 정규화된 음향변이(발화 음성에서 추출된 ΔHz(주파수 변화량), ΔdB(강도 변화량), Δms(발성 길이 변화량)을 정규화한 후, 감정 변화의 방향(증가/감소/유지)을 8단계가 임계 값 기반 이진화를 통해 3비트 감정코드로 변환된다[7][14-15].

이로써, 기존 시각화 방식과 캔들코드의 차별성은 아래와 같다. 기존 방식은 정량 정보는 제공하나 감정 패턴을 명확히 제시하지 못한다.

하지만, 제시한 캔들코드는 증가/감소/기준선 유지의 패턴을 시간축에 따라 직관적으로 이해할 수 있다. 이는 캔들코드는 감정의 시간적 전이 패턴을 직관적으로 표현함으로서 BERT 감정모델의 데이터보호 관점에서 비식별화가 불가능 구조를 극복한 개인 맞춤형 감정 모니터링의 가치를 가진다.

이로써, 향후 제시한 본 프레임워크(BEC, Binary Emotion Code)는 실시간 감정 모니터링 서비스와 멀티모달 연계를 통해 프라이버시 보호형 감정 분석 서비스 플랫폼으로의 발전을 기대한다.

REFERENCES

[1] James A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology,* Vol. 39, No. 6, pp. 1161-1178, 1980.

[2] Thomas A. Ostermann et al, "Associations of Personality, Physical and Mental Health with Voice Range Profiles, " *Journal of Voice,* Vol. 39, no. 3, May 2025.

[3] Liu et al, "EASY: Emotion-Aware Speaker Anonymization," arXiv, 2024.

[4] Yu, L. et al, "Multimodal Sensing-Enabled Large Language Models for Automated Emotional Regulation," *Sensors 2025,* 1 Aug. 2025.

[5] Kwon J.Y, et al, "Privacy-Preserving Real-time Speech Signal Arousal Monitoring Using STT and Candlestick Charts," *2025 World Robot Olympiad Future Scientist Journal,* 2025.

[6] J. Fernandes et al, "Harmonic to Noise Ratio Measurement - Selection of Window and Length," *Procedia Computer Science,* vol. 138, pp. 280 - 285, Jan. 2018.

[7] Kang K.M. et al "Method of emotion monitoring and mental diagnosis of patients by use of emotional and acoustic analysis for dialog language collection," *Korea Patent(10_2024_44533),* 2024.

[8] M. Brockmann,, et al, "Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task," *J Voice,* vol. 25, no. 1, Jan. 2011.

[9] A. De Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America,* vol. 111, no. 4, Apr. 2002.

[10] S. Graf, T. Herbig, M. Buck, and G. Schmidt, "Features for voice activity detection: a comparative analysis," *EURASIP Journal on Advances in Signal Processing,* vol. 2015, no. 1, Nov. 2015.

[11] Sudarshan Pant et al, "Korean Drama Scene Transcript Dataset for Emotion Recognition in Conversations," *IEEE Access,* Vol. 10, , 11 Nov. 2022.

[12] Wu, Y. Mi, Q. Gao, T. "A Comprehensive Review of Multimodal Emotion Recognition: Techniques, Challenges, and Future Directions," Biomimetics 2025.

[13] Chen, J. et al. "Qieemo: Speech Is All You Need," 2025(arXiv), https://doi.org/10.48550/arXiv.2503.22687

[14] Kang M.G., "Sentiment communication system based on multiple multimodal agents," *Korea Patent(10-1652486),* 2016.

[15] Seo, J. et al. "FIDO transaction authentication of emotional reaction for contents in usability evaluation," *Korea Patent(10-2024-0083206),* 2022.

저 자 소 개

강민구(정회원)

1986년 연세대학교 전자공학과 공학사
1989년 연세대학교 전자공학과 공학석사
1994년 연세대학교 전자공학과 공학박사
1985~1987년 삼성전자 통신연구소 연구원
2000년~현재 한신대학교 AI.SW대학 교수

<주관심분야 : 통신네트워크, IoT센서, 스마트미디어 콘텐츠>