Lost in Compression: Who's Heard and Who's Blurred in Digital Voice Communication?

A Technical Report submitted to the Department of Systems and Information Engineering

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Elizabeth J. Recktenwald

Spring, 2025

Technical Project Team Members
Catherine H. Nguyen
Madison C. Sullivan
Lucas Vallarino

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Dr. Matthew Bolton, Department of Systems and Information Engineering

Lost in Compression: Who's Heard and Who's Blurred in Digital Voice Communication?

Elizabeth Recktenwald, Catherine Nguyen, Madison Sullivan, Lucas Vallarino,
Jad Atweh, & Matthew Bolton, *Senior Member IEEE*

Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA, USA

***Abstract*—Digital communication has become essential post-COVID-19, yet speech transmission technologies introduce unintended distortions that may disproportionately affect certain voices. This study investigates whether three widely used compression and decompression algorithms (codecs)—OPUS, Adaptive Multi-Rate (AMR), and CODEC2—introduce structural bias by differentially degrading male and female voices. Using a dataset of 2,953 speakers, we applied 14 audio quality metrics to assess differences in voice degradation across various bitrate settings. Our results reveal consistent, metric-dependent disparities by sex, with OPUS showing a shift in bias at mid-range bitrates, and AMR and CODEC2 exhibiting persistent but variable bias across all settings. Results indicate that female voices experience greater degradation across multiple measures, potentially impacting listener perception and communication clarity. These findings highlight the importance of codec design in equitable voice representation and call for greater attention to inclusion in speech processing systems.**

***Index Terms*—Digital Communication, Bias, CODEC, Voice**

I. Introduction

Digital communication has become central to everyday life, particularly in the post-COVID-19 era. As reliance on platforms like Zoom and Microsoft Teams grows, the technologies used to transmit speech are actively shaping how voices are heard. In this situation, it is critical that all voices be presented equitably to enable equal and effective communication. However, there is growing evidence of bias in digital speech codecs: the algorithms that determine how audio is compressed and decompressed. These algorithms often modify the audio to reduce the amount of data transmitted. In particular, literature suggests that popular codecs may be disproportionately degrading female voices. This phenomenon may not merely be a technical flaw, but a barrier to clear and equitable communication, with implications for professional, educational, and social interactions.

This research specifically investigates how different speech compression algorithms alter the acoustic properties of male and female voices, with a focus on quantifying bias across multiple codecs and bitrates (the parameter that controls the limits on data transmission and thus compression levels). By comparing original and degraded WAV files, key objective metrics were selected to mathematically and statistically evaluate potential distortions in digital audio processing. The findings will contribute to a broader understanding of how speech technology can be improved to ensure more equitable voice representation across digital platforms.

In what follows, the background section provides necessary information about digital voice codecs and the evidence for them having biases. This section also discusses different codecs and objective metrics for evaluating audio. The methods section then details the approach we used to analyze for potential sexual biases in digital audio codecs and bitrates using a large database of human voices. The results section presents the raw findings of our analyses. These findings are further explored in the discussion to provide deeper insight into their implications. Finally, the potential limitations and areas of future work in research are explored followed by the conclusion which summarizes the main takeaways.

II. Background

This section provides background on topics necessary to understand our analyses. This includes a general description of digital voice communication, the evidence for bias in it, the vocal codecs we considered in this research, as well as the metrics used in our analysis.

A. Digital Voice and Evidence of Bias in Codecs

Digital voice communication occurs when analog vocal sound is encoded into a digital signal, transmitted, then decoded back into analog audio, where it can be listened to. For example, Voice over Internet Protocol (VoIP) is a form of communication that transmits voice data over the internet. Due to their ability to provide real-time communication, platforms such as Zoom, Discord, and Microsoft Teams use the VoIP system. There is also digital radio, where digitally encoded audio is transmitted using radio signals [1].

Codecs are integrated into digital voice platforms to process audio from sender to receiver. A codec is an algorithm that enables a message to be transmitted through the encoding and decoding of data. The purpose of encoding and decoding data is to shrink audio files to efficiently transmit over the internet.

Most codecs are lossy, meaning that audio data is eliminated to minimize data transmission. Sound quality is determined by codec settings. Each setting has a bitrate (usually measured in kilobits per second, kbps), which indicates the amount of data transmitted per second. A lower bitrate transmits lower audio quality, which usually means it allows for the loss of more audio data. There are different strategies for eliminating audio data to match bitrates. This can include applying band filters, which constrain the frequencies transmitted by cutting off low and high frequency data. For example, narrowband, which is used at low bitrates and only transmits frequencies from 3 Hz – 4000 Hz. Mediumband transmits frequencies from 3 Hz – 6000 Hz, wideband from 3 Hz – 8000 Hz, super wide from 3 Hz – 12000 Hz, and full from 3 Hz – 20000 Hz. The bitrate used by the codec, and thus the band, is often determined by the bandwidth of the transmission environment. For example, a slow internet connection could result in VoIP using a low bitrate and narrowband [2].

Unfortunately, there is preliminary evidence that codec bitrate and band can distort female voices more than male's. Bolton [3] provided preliminary evidence of sexual bias in the lowest bitrate setting (6 kbps) of the popular OPUS audio codec for multiple objective measures of audio degradation. Similarly, Daum [4] examined the impact of audio artifacts on perceptions of male and female speakers, indicating that the lowest bitrate of the AMR codec (one used in mobile phone communications) negatively affects the intelligibility and perceived quality of female voices. Additionally, Ireland et al. [5] found that lower bitrates in Adaptive Multi-Rate (AMR) compression, such as 4.75 kbps and 5.9 kbps, introduced distortions that could negatively affect the analysis and perception of vowel quality, particularly in female voices. While these studies have established the presence of bias, gaps remain in understanding how specific audio codecs and their bitrate configurations exacerbate or mitigate this issue. This is due to the existing analyses only considering the lowest bitrates of select codecs. As such, existing literature lacks a comprehensive analysis of these widely used codecs across multiple bitrates, a gap this study aimed to address.

B. Digital Voice Codecs

There are many different audio codecs used in digital voice communication. We surveyed various ones based on factors that included popularity, accessibility, and sound-processing capability. We considered what applications each codec is used for, if they are easily obtainable for research, and their bitrate range, respectively. In this research, we considered three popular codecs: OPUS, AMR, and CODEC2. All are open-source and vary by application usage and capability [6].

- OPUS is part of the Web Real-Time Communication standard and is used in common applications such as Discord, WhatsApp Zoom, and Microsoft Teams. It is the most versatile of the three codecs, offering narrowband, mediumband, wideband, super wideband and fullband bitrate settings to handle a wide range of audio qualities. Its bitrate settings range from 6 to 510 kbps. For our analysis, we focused on bitrates from 6 through 48 kbps – the ones used for VoIP [2].
- AMR (Adaptive Multi-Rate Speech Codec) is used for GSM (Global System for Mobile Communications) and UMTS (Universal Mobile Telecommunications Systems) - 3G mobile device communication. It has eight bitrates covering a narrowband range from 4.75 to 12.2 kbps [7].
- CODEC2 is the leading open-source codec used for High Frequency/Very High Frequency radio (for amateur, non-commercial, and emergency applications). Its bitrates cover a narrowband range from 0.7 to 3.2 kbps. CODEC2 is useful for being an open-source voice codec with a very narrowband (low) bitrate [8].

C. Metrics

We surveyed the literature to identify objective metrics that would be appropriate for measuring distortion caused by audio compression in our study. For this, we identified a number of sources that reviewed different options [3], [9], [10]. This ultimately resulted in the metrics shown in Table I. Note that several of these metrics rely on the use of Fourier transforms. Thus, these use standard voice parameters of a 25ms window size, an overlap of 12.5 ms, and 256 samples [6].

One of the 14 metrics, PESQ, operates in two modes: narrowband (8 kHz sampling rate) and wideband (16 kHz sampling rate) [11], [12], [13]. PESQ2 corresponds to narrowband, while PESQ1 corresponds to wideband.

III. Objectives

Existing studies have highlighted the presence of a bias that appears to distort female voices more than male ones in digital voice communication codecs. However, there is a gap in understanding how specific audio algorithms—namely OPUS, AMR, and CODEC2—and their configurations (bitrates) exacerbate or mitigate this issue. This paper aims to fill this gap by quantifying the extent to which bias appears across objective metrics of distortion. To address this, we sought to gather a large database of voice and compress them across the range of bitrates supported by these codecs and see where and how bias is manifested between the sex of the speakers.

IV. Methods

The dataset was selected after searching for a large database of speech, with a diverse range of accents, that included the same spoken phrase in English. The chosen database was the Speech Accent Archive [14]. This included 2,954 unique people, with an approximately equal distribution of male (1,440) and female (1,514) speakers with a variety of ages and language backgrounds. Demographic information for each speaker also included their age, native language, and accent. Each of these original files was represented as a standard, uncompressed, 16-bit WAV file with a sample rate of 44.1 kHz. Each file in this database was encoded/compressed using our identified codecs (OPUS, AMR, and CODEC2) for multiple bitrates. Each file was then losslessly decoded back to a WAV file so that this compressed version could be compared against its original. For AMR and CODEC2, this was done for every supported bitrate: 4.75, 5.15, 5.9, 6.7 7.4, 7.95, 10.2, and 12.2 kbps for AMR, and 0.7, 1.2, 1.3, 1.4, 1.6, 2.4, and 3.2 kbps for CODEC2. This could not be done for OPUS because OPUS supports a continuous range of bitrates from 6 to 510 kbps, where the bitrate influences the signal band. Because we expected to find the most bias in narrowband, but wanted to oversample in that range, while still covering the entire bitrate range. This meant that we selected four bitrates: 6, 6.7, 7.4, and 7.95 kbps that

TABLE I
OVERVIEW OF METRICS FOR AUDIO QUALITY EVALUATION

Metric	**Definition**	**Optimal**
Centroid Difference	The difference in mean frequencies of the original and compressed audio samples, measured in Hz.	Lower
Coherence Value	The degree (magnitude-squared coherence) to which two signals are linearly related in the frequency domain.	Higher
Compression Error	Absolute error between the frequencies of the signal over time for the original and compressed audio files.	Lower
Cross Correlation	A measure of the correlation of sound pressures to show how similar or different the sound pressure is for the original and compressed data.	Higher
Euclidean Distance	The square root of the sum of squared error between the original audio pressure level and the pressure level of the compressed audio.	Lower
MOS	MOS is the outcome of ViSQOL, which is a spectro-temporal (relationship between frequencies and time) measure of the similarity between original and compressed audio data that is translated to a mean objective score between 1–5.	Higher
ODG	Quantifies how the human ear perceives degradation in sound quality between original and processed audio files. 0: no degradation, -4: significant degradation.	Lower
PESQ	Evaluates speech quality using a model of human auditory perception. Lower score, lower quality. Higher score, higher quality. 1.5 = bad, 4.5 = perfect.	Higher
RMSE	A measure of the difference between the original audio pressure level and the pressure level of the compressed audio that calculates the square root of the average of the squared differences between the files.	Lower
Signal Distortion Ratio	Compares the power (intensity or volume) of compressed ("signal") to the original ("noise") in dB.	Higher
Spread Difference	The spectral spread of a sound is a measure of the variance of a signal's frequencies, the "spread" of the sound around its spectral centroid.	Lower
Spectral Entropy	How power is distributed across frequencies and how similar this is across both original and compressed audio samples. This quantifies the irregularity or complexity of a signal based on its frequency content and power distribution.	Higher
STOI	Intelligibility measure that is highly correlated with the intelligibility of degraded speech, e.g., due to added noise.	Higher
Structural Similarity	Measures the similarity between two signals based on their structural information (luminance, contrast, and structure).	Higher

spanned the bitrates to just below the rating (8 kps), where OPUS was guaranteed to operate at mediumband (OPUS may use either narrowband or mediumband from 8 - 10 kbps). We then selected a range of 2 bitrates that were guaranteed to operate at the mediumband (10.2 and 12.2 kbps) and then bitrates that consistently would cover the remainder of the bandwidths in 8 kbps increments up to the top of the VoIP range: 16 kbps to produce wideband, 24 kbps for super wide, and 32, 40, 48 kbps for full.

We developed a MATLAB script that processed each original and corresponding compressed file and computed each of the important metrics we identified in Table I. The results were output in a CSV file for each codec and bitrate.

Next, we wrote a Python script that read each CSV file and ran a Mann-Whitney U non-parametric test to determine if there were significant differences between the sex of the speakers for each metric for each combination of codec and bitrate. A Mann-Whitney U test was used for analysis because of the non-normal distribution and unequal variance of the majority of the data. This test produced corresponding U, p, and r (effect size) statistics and associated median metric values based on sex. The significance level was determined based on a Bonferroni correction that accounted for the multiple comparisons necessary for all 14 metrics. This resulted in an α level of $0.05/14 = 0.00357$. Furthermore, the script produced violin plots for each metric at each bitrate for each codec to visualize the difference between the medians and distributions between females and males. We examined these results to understand whether males or females performed better for metrics where significant differences occurred.

V. RESULTS

Figs. 1 to 3 show metrics that indicate significant differences between male and female voices for every bitrate for OPUS, AMR, and CODEC2, respectively. In these figures, the metrics are presented on either side of a horizontal bitrate axis, with those above the axis favoring females and those below favoring males. Each entry details the metric along with its associated U test statistic and r effect size at each specific bitrate.

Fig. 1 depicts our results for OPUS. This shows that for the lower bitrates (6 kbps to 7.95 kbps), all significant metrics favor males. At 10.2 kbps, there is a switch and an equal or greater number of significant metrics favor females for the rest of the bitrates. The effect size ranges from small (0.07) to moderate (0.3) using standard interpretation heuristics [15].

Fig. 2 shows our results for AMR. The AMR results show that all significant metrics, except Coherence Value, favor males at all bitrates. The effect size ranges from small (0.06) to moderate (0.4).

Fig. 3 reports our results for CODEC2. The results show that for all bitrates, there are more significant metrics that favor males than females. Spectral Entropy, Euclidean Distance, and Coherence Value consistently favor females. The effect size once again ranges from small (0.07) to moderate (0.3).

Notes: (U-value, r-value)
p-value is < 0.001 and is not shown unless 0.001 or more, than indicated in *, **, or ***

* p-value ≈ 0.001
** p-value ≈ 0.002
*** p-value ≈ 0.003

More notes:

PESQ_MOS_1 = PESQ1
PESQ_MOS_2 = PESQ2

6-12k = MOS_2
12.2-48k = MOS_1

sample size - 2 = 2951 DF

					4	4	3	2	
Female	0	0	0	0	**Structural Similarity** (989492, 0.10) **Signal Distortion Ratio** 942801, 0.14) **ODG** (1304191, 0.20) **Cross Correlation** (943044, 0.14)	**Structural Similarity** (1004956, 0.08) **Signal Distortion Ratio** (972918, 0.11) **ODG** (1318512, 0.21) **Cross Correlation** (972879, 0.11)	**Signal Distortion Ratio** (972918, 0.11) **ODG** (1318512, 0.21) *** **Cross Correlation** (972879, 0.11)	**Spread Difference** (1165007, 0.07) * **Spectral Entropy** (1002272, 0.08)	
Bitrate Setting	*6 kbps*	*6.7 kbps*	*7.4 kbps*	*7.95 kbps*	*10.2 kbps*	*12.2 kbps*	*16 kbps*	*24 kbps*	...
Male	**Compression Error** (918342, 0.16) **Euclidean Distance** (839540, 0.23) **ODG** (986117, 0.1) **$PESQ_2$** (662770, 0.12) **RMSE** (838513, 0.23) **Spectral Entropy** (1264887, 0.16) **Structural Similarity** (1239620, 0.14) **Signal Distortion Ratio** (1177757, 0.08)	**Compression Error** (902681, 0.17) **Euclidean Distance** (819241, 0.25) **ODG** (998804, 0.08) **$PESQ_2$** (668694, 0.11) **RMSE** (818237, 0.25) **Spectral Entropy** (1259436, 0.16) **Structural Similarity** (1274642, 0.17) **Signal Distortion Ratio** (1207067, 0.11) **STOI** (1170153, 0.07) **Cross Correlation** (1166467, 0.07)	**Compression Error** (879367, 0.19) **Euclidean Distance** (793169, 0.27) **ODG** (986187, 0.1) **$PESQ_2$** (654082, 0.13) **RMSE** (792225, 0.27) **Spectral Entropy** (1247660, 0.15) **Structural Similarity** (1314783, 0.21) **Signal Distortion Ratio** (1242429, 0.14) **STOI** (1196469, 0.1) **Cross Correlation** (1208521, 0.11)	**Compression Error** (872923, 0.2) **Euclidean Distance** (786167, 0.28) **ODG** (979084, 0.1) **$PESQ_2$** (646941, 0.14) **RMSE** (785209, 0.28) **Spectral Entropy** (1244611, 0.14) **Structural Similarity** (1328184, 0.22) **Signal Distortion Ratio** (1251617, 0.15) **STOI** (1203977, 0.1) **Cross Correlation** (1220285, 0.12)	**Centroid Difference** (760160, 0.3) **$PESQ_2$** (644792, 0.15) **SSdiff** (863529, 0.21) **Spectral Entropy** (1283269, 0.18)	**Centroid Difference** (927393, 0.15) **$PESQ_1$** (633411, 0.16) **SSdiff** (928409, 0.15) **Spectral Entropy** (1292862, 0.19)	**$PESQ_1$** (633411, 0.16) **Spectral Entropy** (1292862, 0.19) *	**$PESQ_1$** (589436.5, 0.22)	
	8	10	10	10	4	4	2	1	

32k Females: SSdiff (1185768, 0.09), SpEntroyCorr (991911, 0.09), CentroidDifference (1188713, 0.09)
40k Females: SSdiff (1186754, 0.09), SpEntroyCorr (993706, 0.09), CentroidDifference (1189260, 0.09)
48k Females: SSdiff (1184124, 0.09), SpEntroyCorr (991962, 0.09), CentroidDifference (1186062, 0.09)
32k Males: PESQ_MOS_1 (623125.5, 0.17)
40k Males: PESQ_MOS_1 (663294, 0.12)
48k Males: PESQ_MOS_1 (694611.5, 0.08)

Fig. 1. Mann-Whitney U statistic results for all considered metrics for the OPUS codec. Each entry reports a metric that showed a significant difference ($p < 0.00357$) between Male and Female voices. Such entries contain the name of the metric and the associated statistic: (Mann-Whitney U statistic, and r rank-biserial correlation effect size). Entries are positioned horizontally to correspond with the listed bitrates. The vertical location of entries indicates if the difference favored female speakers (above) or male ones (below). The numbers above and below each column report how many metrics are on a given side, thus indicating how many metrics favored females or males for a bitrate. The 32 kbps–48 kbps bitrate results had the same metric distribution observed for 24 kbps. The statistics for results at these bitrates appear in the box in the figure's lower right corner.

	1	1	1	1	1	1	1	1
Female	**Mean Coherence** (670626, 0.39)	**Mean Coherence** (659736.5, 0.4)	**Mean Coherence** (691460, 0.37)	**Mean Coherence** (685482, 0.37)	**Mean Coherence** (737212, 0.32)	**Mean Coherence** (712685, 0.35)	**Mean Coherence** (760562, 0.3)	**Mean Coherence** (768634.5, 0.3)
Bitrate Setting	*4.75 kbps*	*5.15 kbps*	*5.9 kbps*	*6.7 kbps*	*7.4 kbps*	*7.95 kbps*	*10.2 kbps*	*12.2 kbps*
Male	**Signal Distortion Ratio** (1269114, 0.16) **Structural Similarity** (1183908, 0.09) **$PESQ_2$** (745939.5, 0.11) **RMSE** (912531, 0.16) **STOI** (1309889, 0.2) **Euclidean Distance** (915725, 0.16) **Compression Error** (999274, 0.08)	**Signal Distortion Ratio** (1205290, 0.11) **Structural Similarity** (1182087, 0.08) **$PESQ_2$** (737168.5, 0.12) **RMSE** (924013, 0.15) **STOI** (1313455, 0.21) **Euclidean Distance** (926998, 0.15) **Compression Error** (1009263, 0.07)	**Signal Distortion Ratio** (1167655, 0.07) **Structural Similarity** (1180969, 0.08) **$PESQ_2$** (748993.5, 0.1) **RMSE** (929851, 0.15) **STOI** (1311171, 0.2) **Euclidean Distance** (932771, 0.14) **Compression Error** (1015045, 0.07) *	**Structural Similarity** (1179739, 0.08) **$PESQ_2$** (748196, 0.1) **RMSE** (932552, 0.15) **STOI** (1310844, 0.2) **Euclidean Distance** (935472, 0.14) **Compression Error** (1018358, 0.07) **	**Structural Similarity** (1179470, 0.08) **$PESQ_2$** (756452.5, 0.09) **RMSE** (934350, 0.14) **STOI** (1310294, 0.2) **Euclidean Distance** (937348, 0.14) **Compression Error** (1020940, 0.06) ***	**Structural Similiraty** (1176867, 0.08) **$PESQ_2$** (769730, 0.08) **RMSE** (936382, 0.14) **STOI** (1314647, 0.21) **Euclidean Distance** (939298, 0.14)	**Structural Similarity** (1181608, 0.08) **$PESQ_2$** (768171.5, 0.08) **RMSE** (936532, 0.14) **STOI** (1284869, 0.18) **Euclidean Distance** (939520, 0.14) **Compression Error** (1009649, 0.07)	**Structural Similarity** (1179880, 0.08) **$PESQ_1$** (777022, 0.07) **RMSE** (912531, 0.16) **STOI** (1282609, 0.18) **Euclidean Distance** (940837, 0.14) **Compression Error** (1011669, 0.07)
	7	7	7	6	6	5	6	6

Fig. 2. Mann-Whitney U statistic results for the AMR codec. See Fig. 1 for figure interpretation.

VI. Discussion

The goal of this study was to examine whether widely used digital voice codecs introduce structural biases in how they compress and transmit speech, particularly between male and female voices. By applying 14 objective metrics to recordings from over 2,900 speakers, we investigated if and how such biases manifest across three popular codecs—OPUS, AMR, and CODEC2—over a wide range of bitrates.

Our findings reveal that all three codecs exhibit measurable bias between male and female voices. At every bitrate and for each codec, we observed significant differences in how speech was processed between sexes. This indicates that codecs may be systematically benefiting one sex over the other. In many cases—particularly at lower bitrates—metrics favored male voices, but this pattern was not universal. Some metrics consistently recommended female voices, and in some instances, the direction of the bias shifted depending with the bitrate and codec. Thus, bias was consistently present, but it did not always point in the same direction or arise for the same reasons.

Among the codecs tested, OPUS offers the broadest range of bitrate settings, spanning narrowband to fullband [2]. As a result, OPUS displayed the greatest variability in bias. A notable inflection point in bias direction appeared around 10.2 kbps, marking the transition from narrowband to mediumband. Below this threshold—specifically between 6 and 7.95 kbps—all significant metrics showed a consistent bias against female voices, with male voices more faithfully preserved. Snow [16], who did not examine the effect of sex, saw similar results, concluding that speech discrimination occurred as frequency filter cutoffs were constrained. Above 10.2 kbps, however, bias persisted but shifted inconsistently between the sexes, depending on the metric. This suggests that as bitrate increases, different elements of the audio signal are either preserved or distorted in ways that could influence perceived fidelity

	3						
	Spectral Entropy (1009091, 0.07)	2	2	2	2	2	2
	Euclidian Distance (1264327, 0.16)	**Euclidian Distance** (1207166, 0.11)	**Euclidian Distance** (1207169, 0.11)	**Euclidian Distance** (1207146, 0.11)	**Euclidian Distance** (1204873, 0.11)	**Euclidian Distance** (1209034, 0.11)	**Euclidian Distance** (1207412, 0.11)
Female	**Mean Coherence** (910770, 0.16)	**Mean Coherence** (922751, 0.15)	**Mean Coherence** (919847, 0.16)	**Mean Coherence** (917347, 0.16)	**Mean Coherence** (915838, 0.16)	**Mean Coherence** (913160, 0.16)	**Mean Coherence** (910104, 0.17)
Bitrate Setting	*0.7 kbps*	*1.2 kbps*	*1.3 kbps*	*1.4 kbps*	*1.6 kbps*	*2.4 kbps*	*3.2 kbps*
Male	**Centroid Difference** (888948, 0.19)	**Centroid Difference** (1001323, 0.08)	**Centroid Difference** (999099, 0.08)	**Centroid Difference** (1003397, 0.08)	**Centroid Difference** (1000111, 0.08)	**Centroid Difference** (1003931, 0.08)	**Centroid Difference** (995090, 0.09)
	Compression Error (990308, 0.09)	**Compression Error** (999000, 0.08)	**Compression Error** (999514, 0.08)	**Compression Error** (998909, 0.08)	**Compression Error** (1001205, 0.08)	**Compression Error** (993389.5, 0.09)	**Compression Error** (988311, 0.09)
	STOI (1211382, 0.11)	**ODG** (942101, 0.14)	**MOS** (1175535, 0.08)	**MOS** (1171594, 0.08)	**MOS** (1170201, 0.07)	**MOS** (1169958, 0.07)	**ODG** (886481.5, 0.19)
	MOS (1165359, 0.07)*	**$PESQ_2$** (804547, 0.26)	**ODG** (974600, 0.11)	**ODG** (971852.5, 0.11)	**ODG** (926257, 0.15)	**ODG** (989073, 0.09)	**$PESQ_2$** (806400.5, 0.26)
	$PESQ_2$ (910991.5, 0.16)	**RMSE** (919683, 0.16)	**$PESQ_2$** (793858, 0.27)	**$PESQ_2$** (783749, 0.28)	**$PESQ_2$** (779003.5, 0.29)	**$PESQ_2$** (763355.5, 0.30)	**RMSE** (907045, 0.17)
	RMSE (907706, 0.17)	**Spread Difference** (1012303, 0.07)	**RMSE** (916889, 0.16)	**RMSE** (913885, 0.16)	**RMSE** (912425, .16)	**RMSE** (909564, 0.17)	**Spread Difference** (1019192, 0.07)**
	Spread Difference (935817, 0.14)	**Structural Similarity** (1241272, 0.14)	**Structural Similarity** (1241190, 0.14)	**Structural Similarity** (1240817, 0.14)	**Structural Similarity** (1241049, 0.14)	**Structural Similarity** (1241481, 0.14)	**Structural Similarity** (1241751, 0.14)
	Structural Similarity (1241092, 0.14)	**Signal Distortion Ratio** (1246434, 0.14)	**Signal Distortion Ratio** (1209126, 0.11)	**Signal Distortion Ratio** (1229548, 0.13)	**Signal Distortion Ratio** (1222560, 0.12)	**Signal Distortion Ratio** (1253131, 0.15)	**Signal Distortion Ratio** (1314872, 0.21)
	Signal Distortion Ratio (1266278, 0.16)	8	8	8	8	8	8
	9						

Fig. 3. Mann-Whitney U statistic results for CODEC2. See Fig. 1 for figure interpretation.

differently for male and female voices.

In contrast, AMR and CODEC2 operate exclusively within the narrowband and showed no such inflection point. Instead, both codecs displayed persistent but shifting biases across their respective bitrates. This suggests that, for these systems, bias is more closely tied to how specific acoustic features are modeled and compressed than to simple bitrate thresholds. It is also important to note that although both AMR and CODEC2 operate within the narrowband, their bitrate ranges do not overlap. Consequently, their outputs cannot be directly compared to assess speech intelligibility or distortion at the same bitrate. Nevertheless, examining how specific metrics behave across these two codecs offers meaningful insight into the nature of bias under narrowband constraints. Two such metrics—mean coherence and Euclidean distance—reveal distinct yet informative patterns.

Mean coherence, which reflects the temporal synchronization and structural alignment of frequency components, consistently favored female voices across all bitrates in both AMR and CODEC2. This aligns with previous work suggesting that female speech can exhibit higher harmonic regularity and greater periodicity, things more easily preserved under narrowband compression [17]. In contrast, the lower pitch and broader spectral content of male voices, while falling into the frequency range preserved in the narrowband, may be more susceptible to distortion due to their structural complexity and codec sensitivity to low-frequency content [18].

These findings are particularly important in the context of how these codecs are used. AMR, especially in the narrowband, is optimized for speech intelligibility under bandwidth constraints and is widely used in mobile communication [19]. The preservation of coherence across all bitrates suggests that although high-frequency components of female voices may be attenuated, the structural integrity of the remaining signal is relatively intact, supporting consistent vocal clarity. This is also critical when evaluating CODEC2, which was designed for low-bitrate radio communications [20].

Euclidean distance, however, diverges in how it reflects bias between these two systems. In AMR, which is designed for mobile voice communication from 4.75 to 12.2 kbps, male voices tend to show lower Euclidean distances. This indicates that the overall structure of the original signal is preserved more faithfully for male speakers. This may be due to AMR's use of Algebraic Code-Excited Linear Prediction (ACELP), which prioritizes preserving the spectral envelope of low- to mid-frequency sounds [7] that are more typical of male speech.

Conversely, CODEC2, which is designed for ultra-low bitrate transmission of 0.7–3.2 kbps in radio communications, Euclidean distance consistently favored female voices. CODEC2's reliance on harmonic sinusoidal modeling allows it to efficiently encode periodic signals with fewer parameters—a structure more representative of higher-pitched, harmonically rich female voices [21]. As a result, the compression process introduces less deviation from the original signal for female speakers, leading to lower Euclidean distances.

These contrasting patterns underscore how codec architecture interacts with vocal characteristics to produce different kinds of bias, even when operating under similar bandwidth constraints. Depending on which aspect of the signal is prioritized (frequency, power, etc.), bias may appear to impact one sex more than the other. Such variability demonstrates that no single measure can fully capture the nuanced ways in which compression algorithms shape voice fidelity. Although both codecs generally favored male voices across several metrics, these exceptions highlight the need to consider what dimensions of vocal communication are most relevant in a given context. For users and developers alike, identifying which aspects of the signal are most critical to clarity, intelligibility, and/or perceptual similarity could inform which codecs and settings are most appropriate for minimizing bias [22].

VII. Limitations and Future Work

While this study provides strong evidence of sex-based bias in digital voice compression, several limitations should be considered when interpreting the results. First, the analysis relied exclusively on objective metrics to assess audio quality.

While these measures are widely used and offer consistency [3], they may not fully capture how listeners perceive audio degradation in real-world contexts. Without subjective listening tests, it remains uncertain whether the detected differences are meaningful to human perception.

Future work could conduct granular, metric-level analyses across specific bitrates within each codec. Rather than assessing codec performance generally, studies should break down how individual metrics behave at each bitrate and whether certain distortions consistently affect one sex more than the other. This approach could uncover important thresholds or transitions where bias becomes most pronounced. Such an analysis would offer actionable insights for codec design and fairness. Future work could also aim to determine whether similar inflection points exist in other mediumband to wideband codecs.

By conducting an analysis across codecs by bitrates and metrics, it is evident that bias exists in both sexes. This has important implications for codec audio compression and design work. Assessing audio metrics through important acoustic features, such as frequency and power, highlights similarities and differences in how codecs process sound. A comprehensive analysis of these features contributes to a better understanding of how structural biases can be addressed to improve sound quality and codec usability for all.

VIII. Conclusion

This study aimed to identify biases in how digital codecs process male and female voices, with particular emphasis on the discrepancies at lower bitrates where distortion is most likely to occur. Our results showed that such biases are not only present across all three codecs analyzed (OPUS, AMR, and CODEC2), but also vary depending on the specific audio quality metrics used. This has practical consequences in contexts where voice clarity, authority, and intelligibility are crucial, such as telecommunications and automated speech recognition. Moving forward, codec developers should consider technological solutions, such as pitch-shifting algorithms and alternative band filtering methods, to mitigate these biases [23]. In an increasingly digital and voice-mediated world, ensuring that all voices are transmitted with equal clarity and fidelity becomes essential for building more inclusive and representative audio technologies.

Acknowledgment

The authors would like to thank Dr. Birgit Pop and Alexander Daum of the Fraunhofer-Institut für Integrierte Schaltungen IIS for their help and support on this project.

References

[1] H. P. Singh, S. Singh, J. Singh, and S. Khan, "Voip: State of art for global connectivity—a critical review," *Journal of Network and Computer Applications*, vol. 37, pp. 365–379, 2014.

[2] Xiph.Org Foundation, "Opus recommended settings," 2023, accessed: 2025-04-06. [Online]. Available: https://wiki.xiph.org/Opus_Recommended_Settings

[3] M. L. Bolton, "Preliminary Evidence of Sexual Bias in Voice over Internet Protocol Audio Compression," in *International Conference on Human-Computer Interaction*. Springer, 2022, pp. 227–237.

[4] A. Daum, "Impact of audio artifacts on perceptions of male and female speakers," Bachelor's thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2023.

[5] D. Ireland, C. Knuepffer, and S. J. McBride, "Adaptive multi-rate compression effects on vowel analysis," *Frontiers in Bioengineering and Biotechnology*, vol. 3, p. 118, 2015.

[6] K. K. Paliwal, J. G. Lyons, and K. K. Wójcicki, "Preference for 20-40 ms window duration in speech analysis," in *2010 4th International Conference on Signal Processing and Communication Systems*, 2010, pp. 1–4.

[7] B. J. Borgstrom, M. van der Schaar, and A. Alwan, "Rate allocation for noncollaborative multiuser speech communication systems based on bargaining theory," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1156–1166, 2007.

[8] D. S. Sunder and R. K. Kushwaha, "Evaluation of narrow band speech codecs for ubiquitous speech collection and analysis systems," in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 93–98.

[9] L. Almér, "Evaluation of the Perceived Speech Quality for G729D and Opus: With Different Network Scenarios and an Implemented VoIP Application," 2022.

[10] I. Siegert, A. F. Lotz, L. L. Duong, and A. Wendemuth, "Measuring the impact of audio compression on the spectral quality of speech data," *Proceedings of the European Symposium on Speech Processing*, pp. 229–236, 2016, accessed: 2025-04-06. [Online]. Available: https://www.essv.de/pdf/2016_229_236.pdf

[11] International Telecommunication Union., "ITU-T Recommendation P.862: Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs," International Telecommunication Union, Tech. Rep., 2001. [Online]. Available: https://www.itu.int/rec/T-REC-P.862-200102-I/en

[12] International Telecommunication Union, "ITU-T Recommendation P.800.1: Mean Opinion Score (MOS) terminology," International Telecommunication Union, Tech. Rep. P.800.1, 2016. [Online]. Available: https://www.itu.int/rec/T-REC-P.800.1/en

[13] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 2, 2001, pp. 749–752 vol.2.

[14] S. H. Weinberger and S. A. Kunath, "The speech accent archive: Towards a typology of english accents," *Language & Computers*, 2011.

[15] D. C. Funder and D. J. Ozer, "Evaluating effect size in psychological research: Sense and nonsense," *Advances in Methods and Practices in Psychological Science*, vol. 2, no. 2, pp. 156–168, 2019.

[16] W. B. Snow, "Audible frequency ranges of music, speech and noise," *The Bell System Technical Journal*, vol. 10, no. 4, pp. 616–627, 1931.

[17] S. Gunasekaran and K. Revathy, "Spectral fluctuation analysis for audio compression using adaptive wavelet decomposition," in *Information Processing and Management*, ser. Communications in Computer and Information Science, V. Das, R. Vijayakumar, and K. Sridhar, Eds. Springer, 2010, vol. 70, pp. 331–338.

[18] A. Bailey and M. D. Plumbley, "Gender bias in depression detection using audio features," *arXiv preprint arXiv:2010.15120*, 2020. [Online]. Available: https://arxiv.org/abs/2010.15120

[19] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (amr-wb)," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 8, pp. 620–636, 2002.

[20] D. G. Rowe, "Codec 2 algorithm description," 2025, accessed: 2025-04-08. [Online]. Available: https://github.com/drowe67/codec2/blob/main/doc/codec2.pdf

[21] P. Jamieson, S. S. Kumar, J. A. M. Nacif, and R. Ferreira, "Analyzing a low-bit rate audio codec - codec2 - on an fpga," in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2021, pp. 1486–1492.

[22] S. Zielinski, "On some biases encountered in modern audio quality listening tests (part 2): Selected graphical examples and discussion," *Journal of the Audio Engineering Society*, vol. 64, no. 1/2, pp. 55–74, 2016.

[23] S. Das, T. Bäckström, and G. Fuchs, "Fundamental frequency model for postfiltering at low bitrates in a transform-domain speech and audio codec," in *Proc. Interspeech 2020*. ISCA, 2020, pp. 2912–2916.