Oct 29, 2025

Characterizing Changes in the Female Voice Across the Last Century using Evidence from the Silver Screen: Protocol for a Methodological Overview

DOI
https://dx.doi.org/10.17504/protocols.io.5jyl8884rl2w/v1

Kelvin Tran[1,2], David Ingram[1,2], Julie Liss[1,2], Visar Berisha[1,2,3]

[1]Arizona State University; [2]College of Health Solutions; [3]Fulton Schools of Engineering

Arizona State University

K Kelvin Tran

DOI: https://dx.doi.org/10.17504/protocols.io.5jyl8884rl2w/v1

Protocol Citation: Kelvin Tran, David Ingram, Julie Liss, Visar Berisha 2025. Characterizing Changes in the Female Voice Across the Last Century using Evidence from the Silver Screen: Protocol for a Methodological Overview. **protocols.io** https://dx.doi.org/10.17504/protocols.io.5jyl8884rl2w/v1

Protocol status: Working
This protocol was used in our longitudinal analysis of the female voice over the last century, using film and media sources of approximately 800 actresses to characterize changes in pitch and speaking rate.

Created: October 28, 2025

Last Modified: October 29, 2025

Protocol Integer ID: 230928

Keywords: Speech, Audio Analysis, Audio Recording, Voice Analysis, Acoustic Analysis, Voice Data, Speaker, Recording, Speech Sample Collection, Speech Database, Protocol, Voice Sample, Utterance, Conversational Speech, longitudinal archive of female speech, changes in the female voice, female speech, speech clip, fidelity speech sample, second speech sample, female voice, longitudinal archive, evidence from the silver screen, interview, featuring actress, wavesurfer

Abstract

This protocol follows a study on assembling a longitudinal archive of female speech spanning from 1929 to 2021. High-fidelity speech samples were extracted from films, television shows, and interviews featuring actresses aged 15 to 25 years old, with a second speech sample extracted approximately 20 years later. The protocol describes how speech clips were pre-processed using TMPGEnc and GoldWave, and further standardized prior to analysis in WaveSurfer, as well as how metadata was collected.

Image Attribution

Figure 1. Statistics Pane & Properties Dialogue of an Actress Speech Sample using the WaveSurfer Plugin API for WaveSurfer v 1.8.8p4 Stable Release

Guidelines

Software that may be needed for reproducing this protocol is described below.

Materials

GoldWave
TMPGExpress or TMPGEnc
WaveSurfer v1.8.8p4

Safety warnings

The version of software used in this procedure is listed throughout the documentation. The user should be aware that when attempting to follow the protocol using different speech analysis or audio editing software that results may vary slightly due to software differences.

Inclusion and Exclusion Criteria

1 Speech samples were included if they met all of the following criteria:

- Speaker: Female actress appearing in a commercially released film, television program, or interview in archive, which included movies between 1920 and present day.
- Age at recording: Between 15 and 25 years for the primary decade sample; a second sample, approximately 20 years later, was included when available.
- Speech content: Conversational English dialogue in the actress's natural speaking voice (non-singing, non-dramatic delivery).
- Audio quality: Minimal background noise, music, or strong emotional affect; at least 15 seconds of continuous speech.
- Data availability: Sufficient metadata available to determine age, height, and birthplace (for dialect classification).

2 Observations were excluded from analysis if they met any of the following conditions:

- Unreliable acoustic estimates: Fundamental frequency (F_0) extraction errors due to noise, clipping, or tracking failures that could not be corrected through manual verification.
- Insufficient duration: Speech segments shorter than 15 seconds after trimming pauses and non-speech intervals.
- Missing metadata: Incomplete or unverifiable demographic variables (e.g., unknown birth year, height, or birthplace) required for model covariates.
- Non-representative vocal style: Speech dominated by extreme emotional prosody, shouting, whispering, or stylized performance inconsistent with natural dialogue.

Voice Sample Selection Criteria

3

- Speech samples were selected from conversational speech between a given actress and a co-star, when applicable.
- Recordings were extracted from scenes featuring the actress's natural speaking voice, with minimal background noise, music, or emotional intensity.
- Only clips with at least 15 seconds of continuous speech were retained for analysis.

Sampling Across Timepoints

4 Two primary timepoints were defined for each actress:

- Between 15-25 years of age, and
- Approximately 20 years from the initial time point

Utterance Quality

5 Approximately 5–10 utterances per actress were selected to reduce variability introduced by noise, recording inconsistencies, or within-film context differences.

Initial Pre-Processing

6 Film clips were processed using TMPGEnc (also known as TMPGExpress Video Mastering Works).

Software

TMPGExpress	NAME
Pegasys	DEVELOPER
https://tmpgenc.pegasys-inc.com/en/product/tvmw7.html	REPOSITORY
https://tmpgenc.pegasys-inc.com/en/product/tvmw7.html	SOURCE LINK

7 Audio tracks were standardized to 16-bit with a sampling rate of 44.1kHz.

8 Once encoded, actress speech segments were extracted using GoldWave, where fine trimming and analysis of speech was guided by real-time spectrograms and the corresponding waveforms in order to isolate the clean speech signal.

Software	
GoldWave	NAME
GoldWave Inc.	DEVELOPER
https://goldwave.com/release.php	REPOSITORY
https://goldwave.com/release.php	SOURCE LINK

Acoustic Measure Extraction

9 WaveSurfer v 1.8.8p4 Stable Release was used for acoustic measure extraction.

Software	
WaveSurfer	NAME
Center for Speech Technology, KTH Royal Institute of Technology, Karolinska Institutet and Stockholm University	DEVELOPER
https://sourceforge.net/projects/wavesurfer/files/wavesurfer/1.8.8p4/	REPOSITORY
https://sourceforge.net/projects/wavesurfer/files/wavesurfer/1.8.8p4/	SOURCE LINK

10 Speaking Rate (syllables/second) was computed by dividing the transcribed syllable count by the total speech duration in seconds.

11 Mean fundamental frequency (F_0) was extracted using WaveSurfer's Statistics Function as shown in the figure below, following procedures described in documentation from Li et al., 2021, and Sjölander & Beskow, 2000.

12 F_0 contours were manually verified against spectrograms for tracking accuracy.

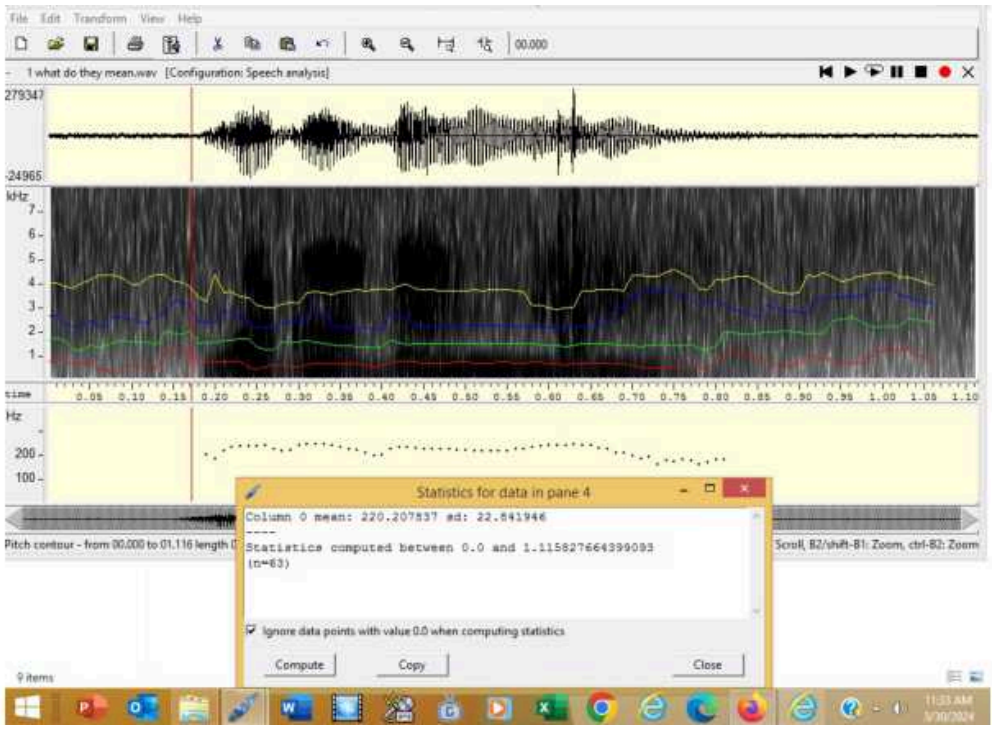

Figure 1. Statistics Pane & Properties Dialogue of an Actress Speech Sample using the WaveSurfer Plugin API for WaveSurfer v 1.8.8p4 Stable Release

Metadata Documentation

13 The type of source media (Film, TV, Online Video, Interview), and year of recording were logged for each extracted sample.

- Actress age at the time of recording was calculated using the actress's recorded birth year against the year of release for the respective source media.
- The actress's height in inches and birthplace were extracted from IMDB biographies.

14 Dialect Regions were classified using the actress's birthplace and categorized according to the TIMIT corpus dialect regions with the following modifications:

- Washington, D.C., was added to the North Midland region
- Hawaii was classified as a separate group
- An additional category was created for a small subset of actresses born outside of the United States

Protocol references

Li, G., Hou, Q., Zhang, C., Jiang, Z., & Gong, S. (2021). Acoustic parameters for the evaluation of voice quality in patients with voice disorders. *Annals of Palliative Medicine, 10*(1), 13036-13136.

Sjölander, K., & Beskow, J. (2000). Wavesurfer-an open source speech tool. In *Sixth International Conference on Spoken Language Processing*.