

Vocal effort modification for singing synthesis

Olivier Perrotin, Christophe d'Alessandro

LIMSI, CNRS, Université Paris-Saclay

Motivations

- Vocal effort
 - Perceived power of the voice
 - Change of loudness + change of spectral properties
 - In speech: modified for intelligibility (increased distance, background noise, etc.)
 - In singing: modified for aesthetic purposes

Motivations

- Vocal effort
 - Perceived power of the voice
 - Change of loudness + change of spectral properties
 - In speech: modified for intelligibility (increased distance, background noise, etc.)
 - In singing: modified for aesthetic purposes
- In singing synthesis
 - Several methods use real voice samples (concatenative synthesis, voice transformation, etc.)
 - Few vocal effort levels are recorded
 - Vocal effort modification techniques provide a way to synthesize a continuous variation of vocal effort

State of the art

- Spectral morphing
 - Extract spectral envelopes from signals with low and high vocal effort
 - Apply a weighted average spectral envelope to one of the signal to synthesize intermediate vocal efforts

State of the art

- Spectral morphing
 - Extract spectral envelopes from signals with low and high vocal effort
 - Apply a weighted average spectral envelope to one of the signal to synthesize intermediate vocal efforts
- Spectral modeling
 - Analyze spectral features responsible for the perception of vocal effort
 - Apply spectral transformations on a signal based on the analysis
 - Decrease vocal effort is easily made by attenuating high frequencies in the signal spectrum
 - Increase effort is harder as it requires to generate high frequencies

State of the art

- Spectral morphing
 - Extract spectral envelopes from signals with low and high vocal effort
 - Apply a weighted average spectral envelope to one of the signal to synthesize intermediate vocal efforts
 - Spectral modeling
 - Analyze spectral features responsible for the perception of vocal effort
 - Apply spectral transformations on a signal based on the analysis
 - Decrease vocal effort is easily made by attenuating high frequencies in the signal spectrum
 - Increase effort is harder as it requires to generate high frequencies
- ➔ We propose a **spectral modeling method to increase vocal effort**

Outline

- Vocal effort modification system
 - Harmonic enrichment
 - Spectral envelope modification
- Evaluation

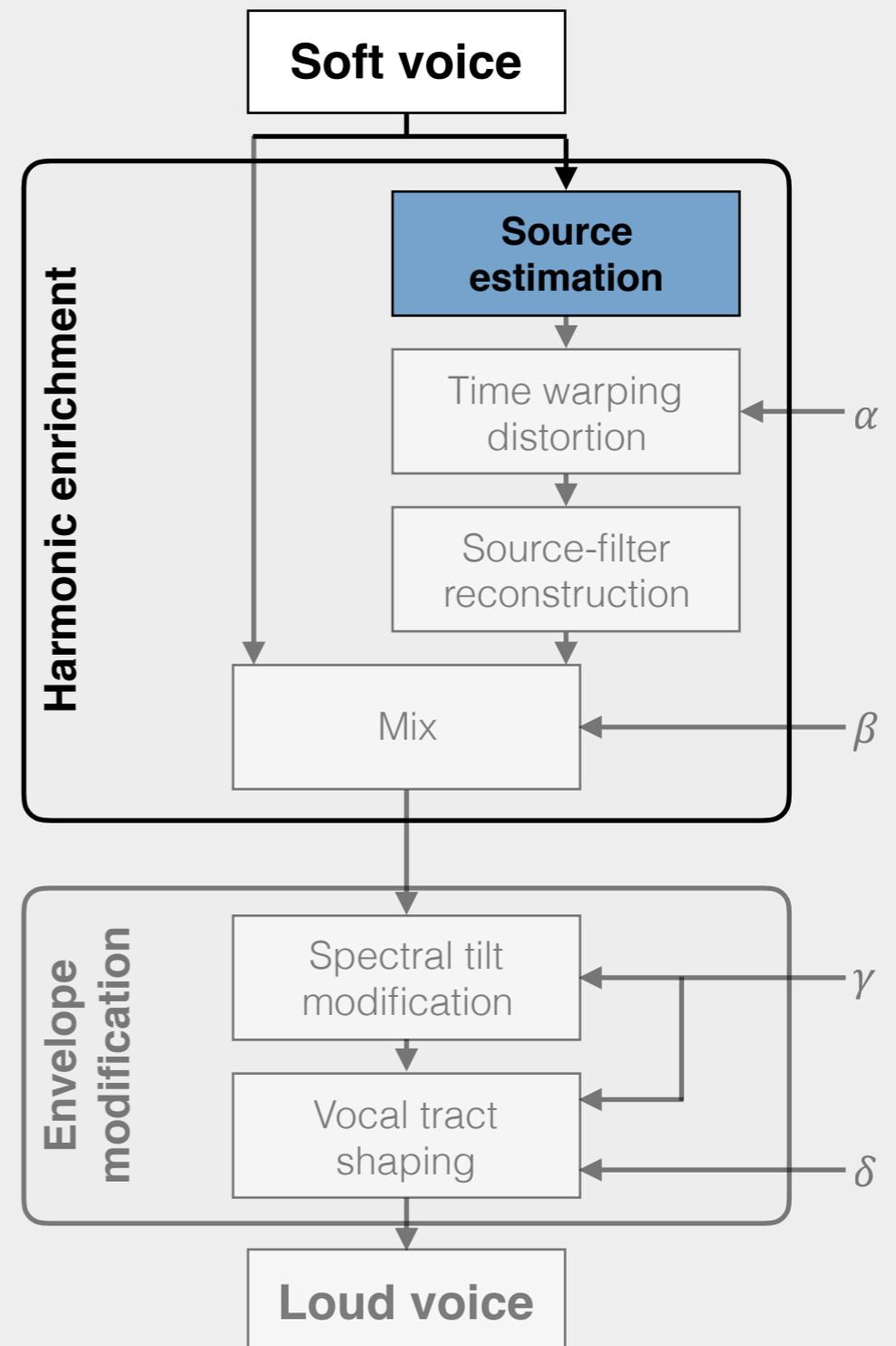
Harmonic enrichment

- Increase of vocal effort produced by a more abrupt closure of the vocal folds
- Sharper notches in the glottal flow variation (derivative)
- Create higher frequencies
- ➔ Harmonic enrichment is made with time distortion to increase the glottal flow derivative sharpness

Harmonic enrichment

Source estimation

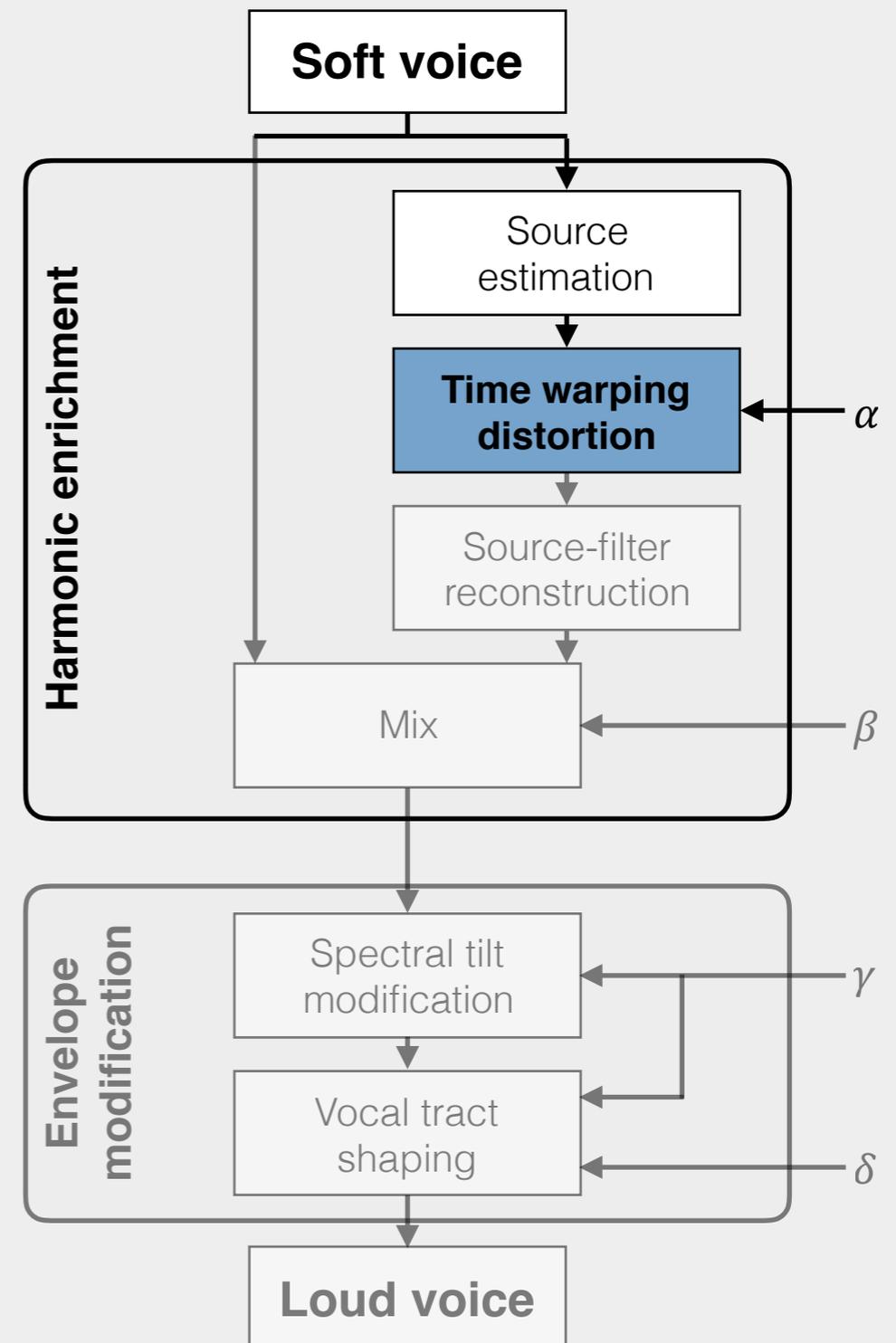
- Get a rough estimation of the glottal flow derivative (source)
 - Attenuate the vocal tract contribution
 - Bandpass filtering [$0.5 F_0 - 1.2 F_0$]



Harmonic enrichment

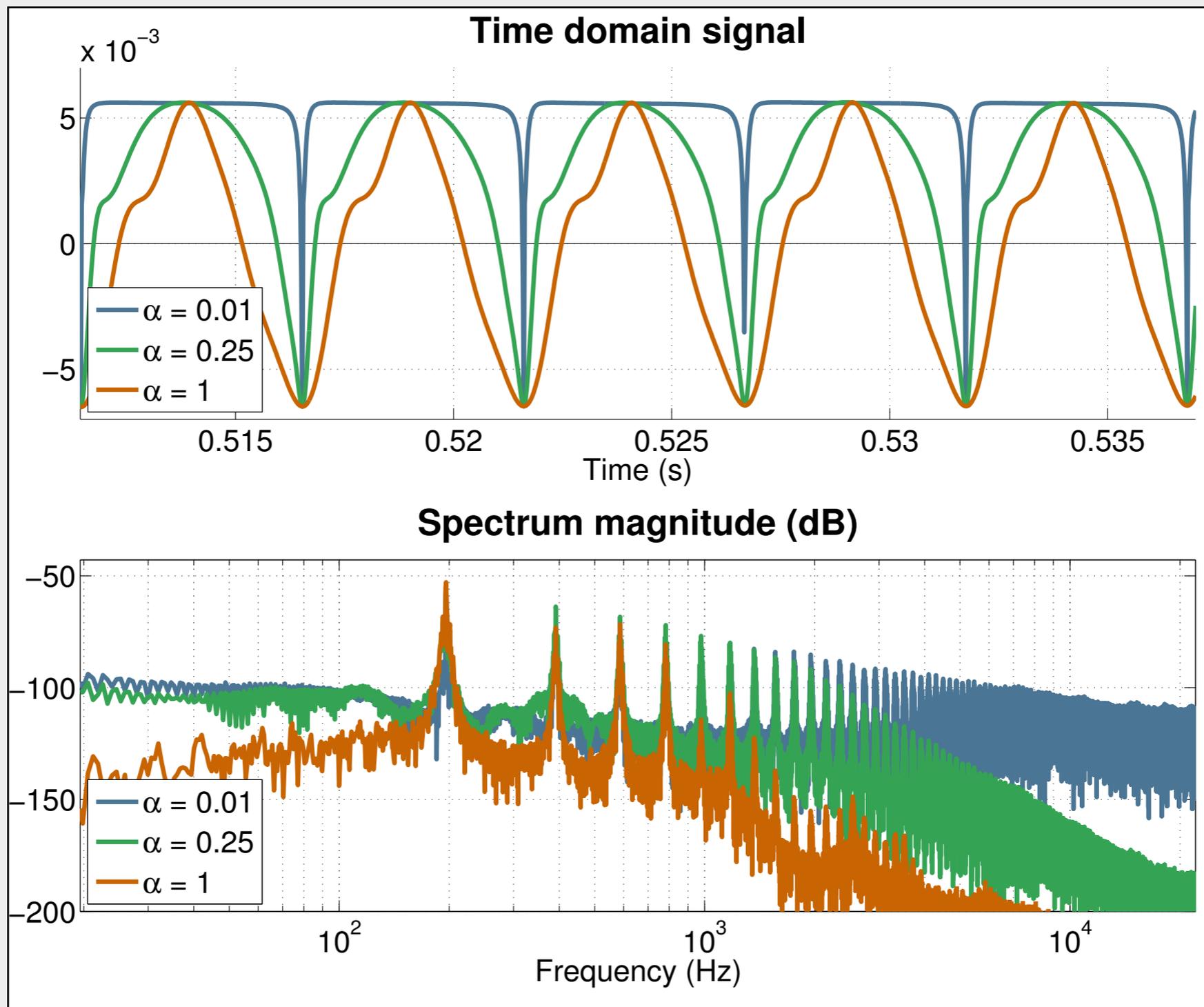
Time warping distortion

- Time distortion
 - Compress the glottal flow derivative around notches
 - Expand the signal in-between
- ➔ Identify notches
- ➔ Apply warping function around each notch with a coefficient α



Harmonic enrichment

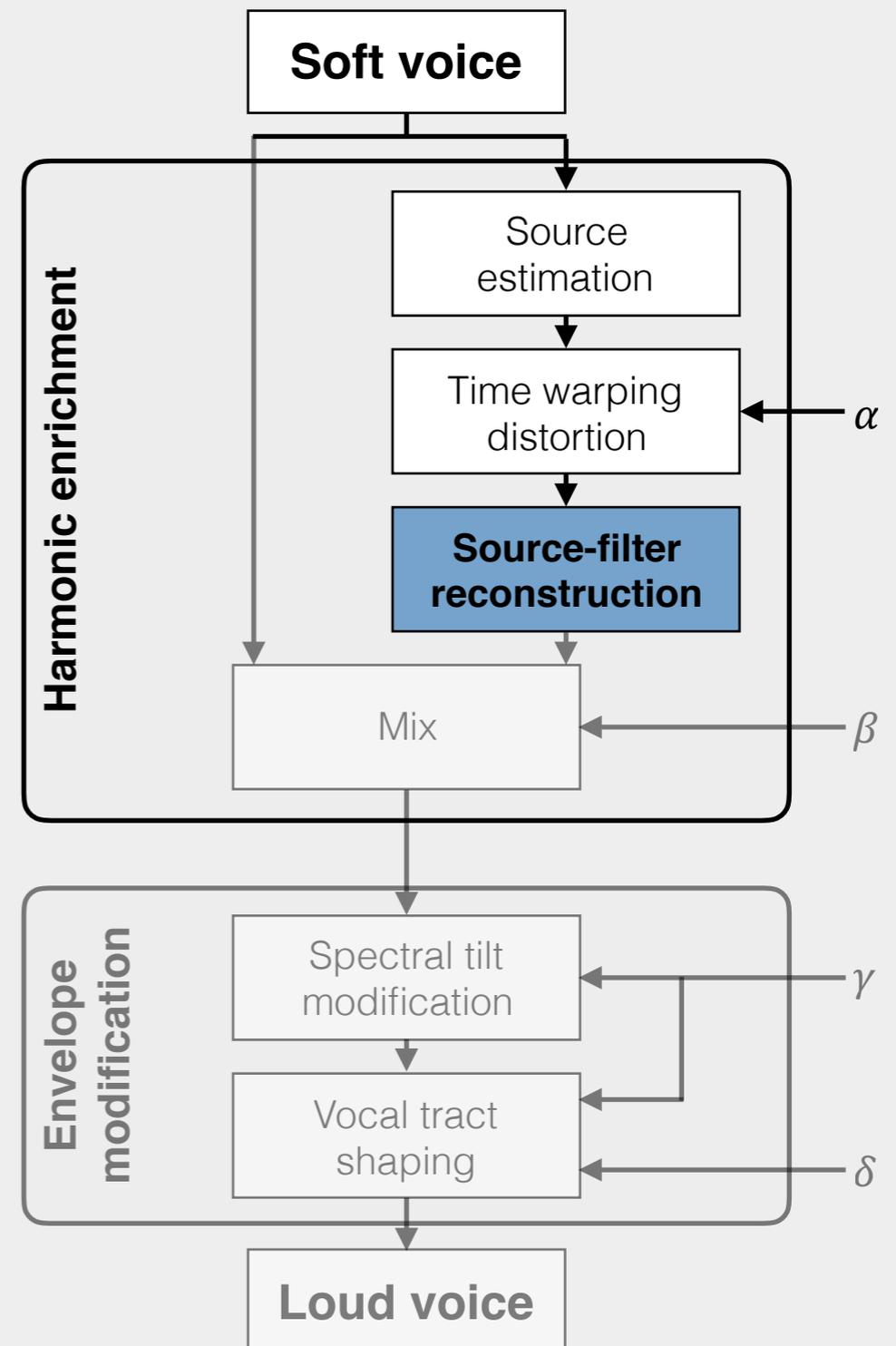
Time warping distortion



Harmonic enrichment

Source-filter reconstruction

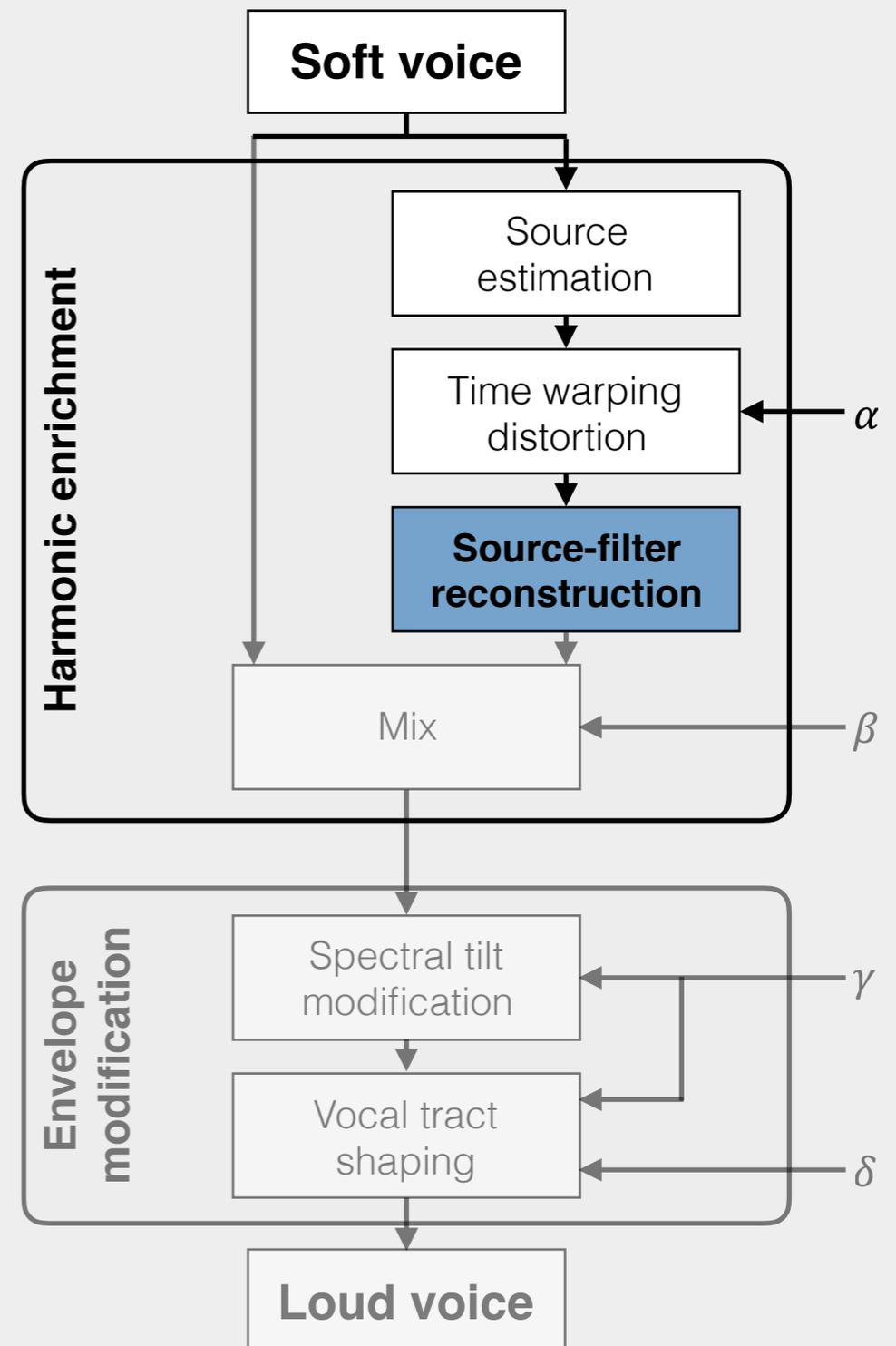
- Reintroduce the vocal tract contribution on the distorted signal
 - Estimate the original signal's spectral envelope
 - Apply it to the distorted signal



Harmonic enrichment

Source-filter reconstruction

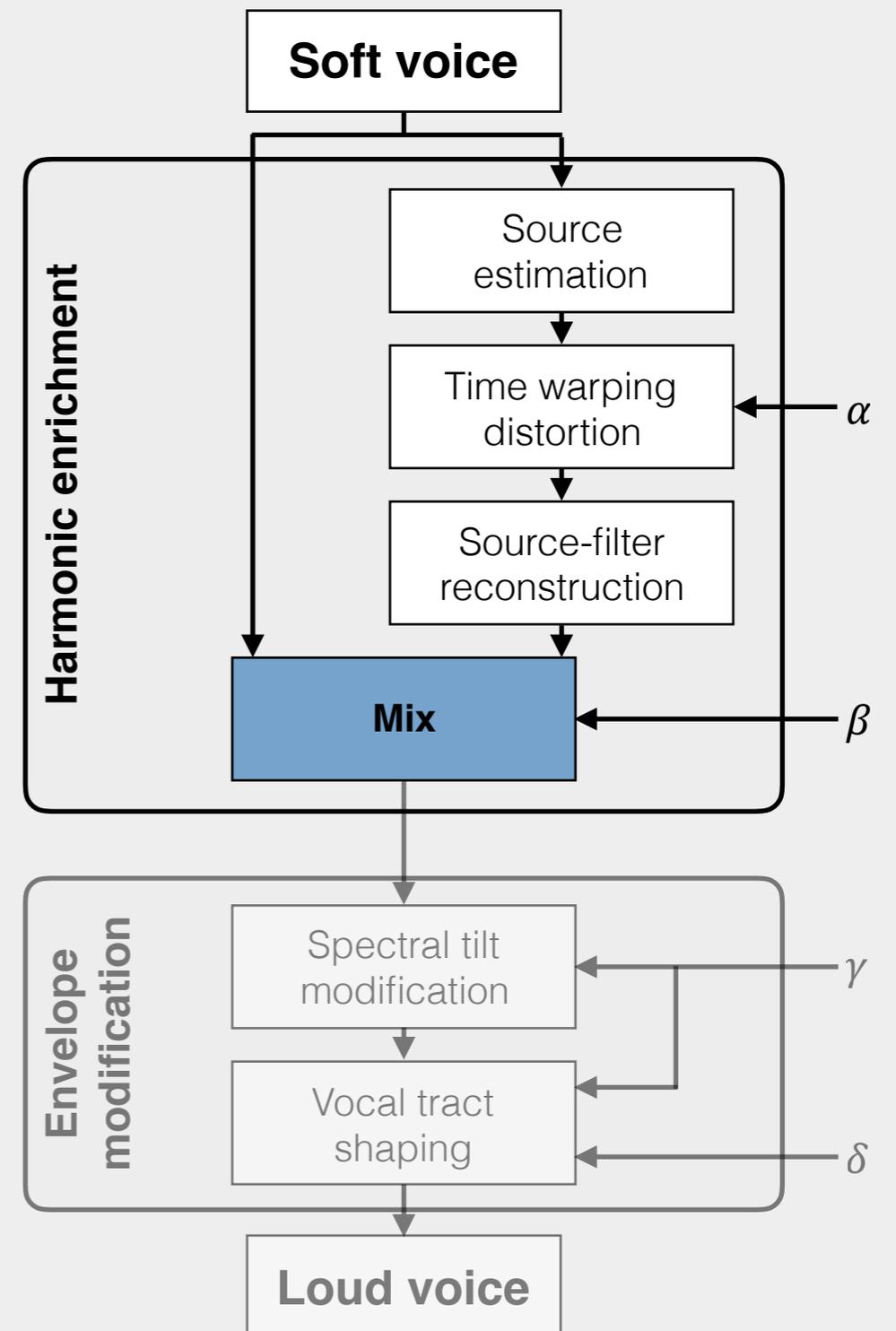
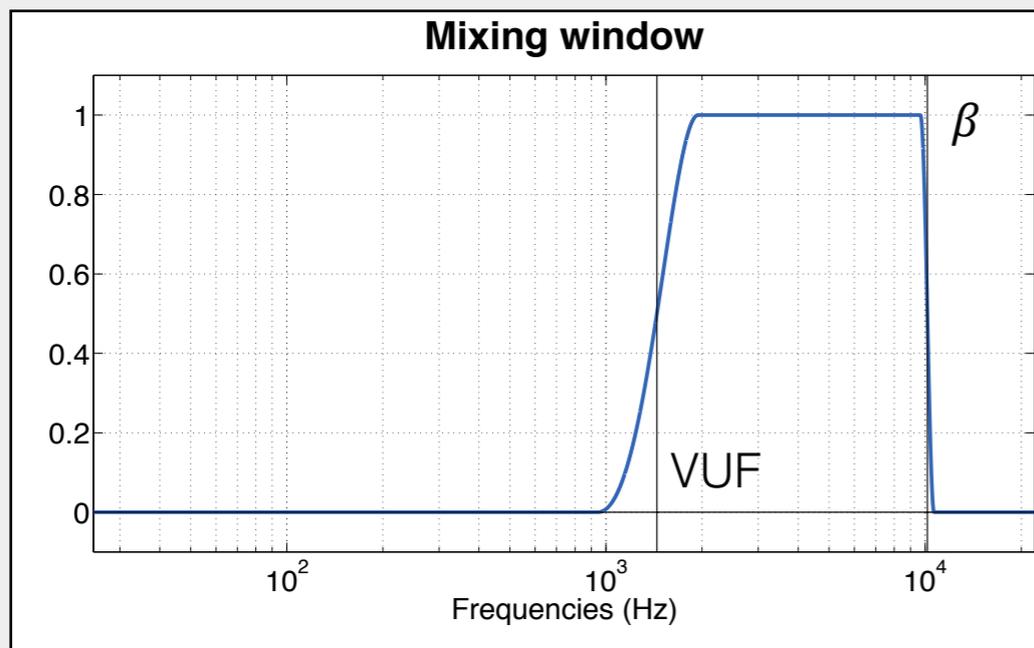
- Reintroduce the vocal tract contribution on the distorted signal
 - Estimate the original signal's spectral envelope
 - Apply it to the distorted signal
- High amount of distortion introduces artifacts in the signal



Harmonic enrichment

Mixing

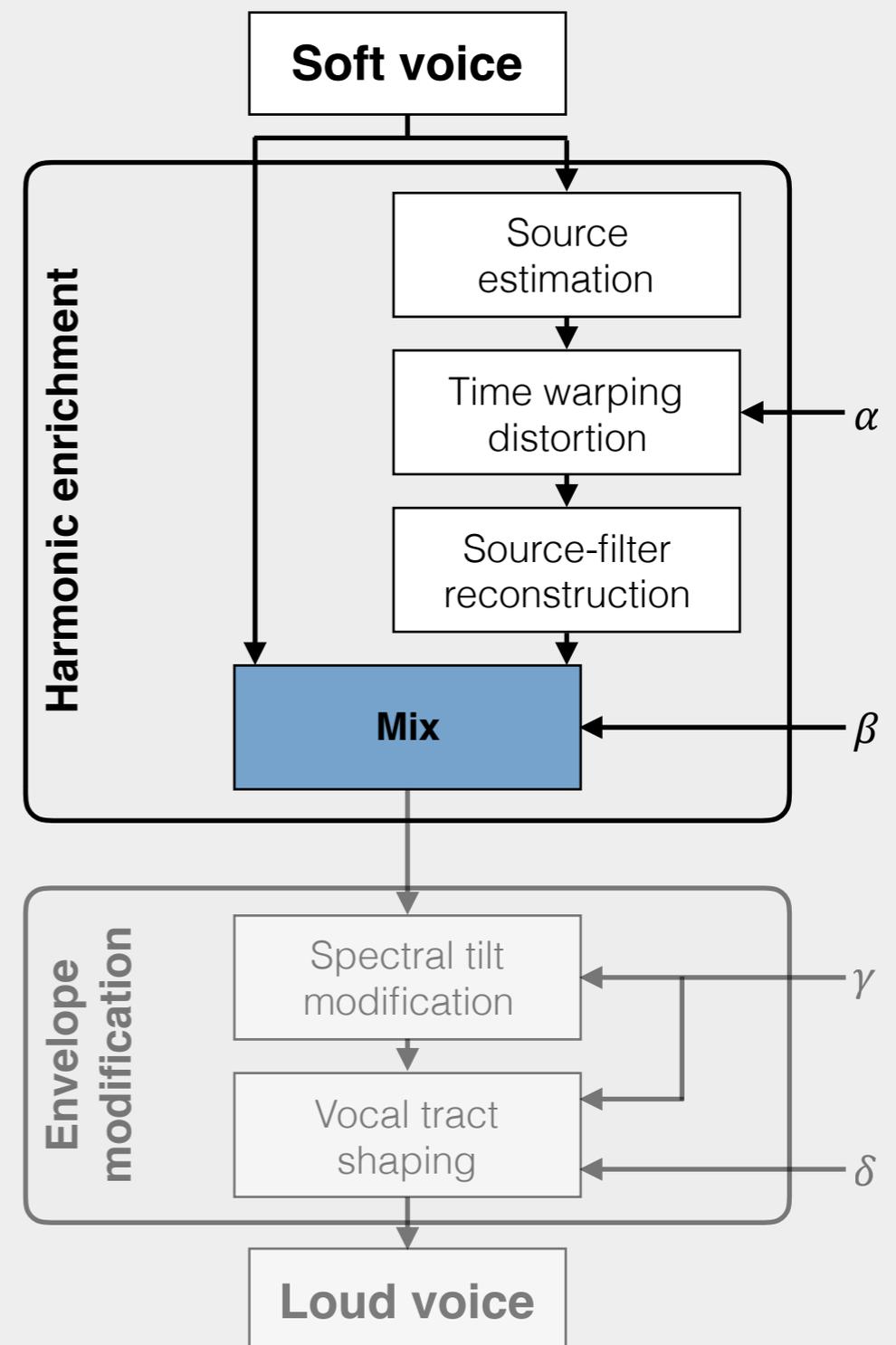
- Introduce new harmonics only where they are missing
 - Design a frequency window on a specific band [Voiced Unvoiced Frequency-10kHz]
 - Mix the original and distorted signals' spectra on this band with a gain β



Harmonic enrichment

Mixing

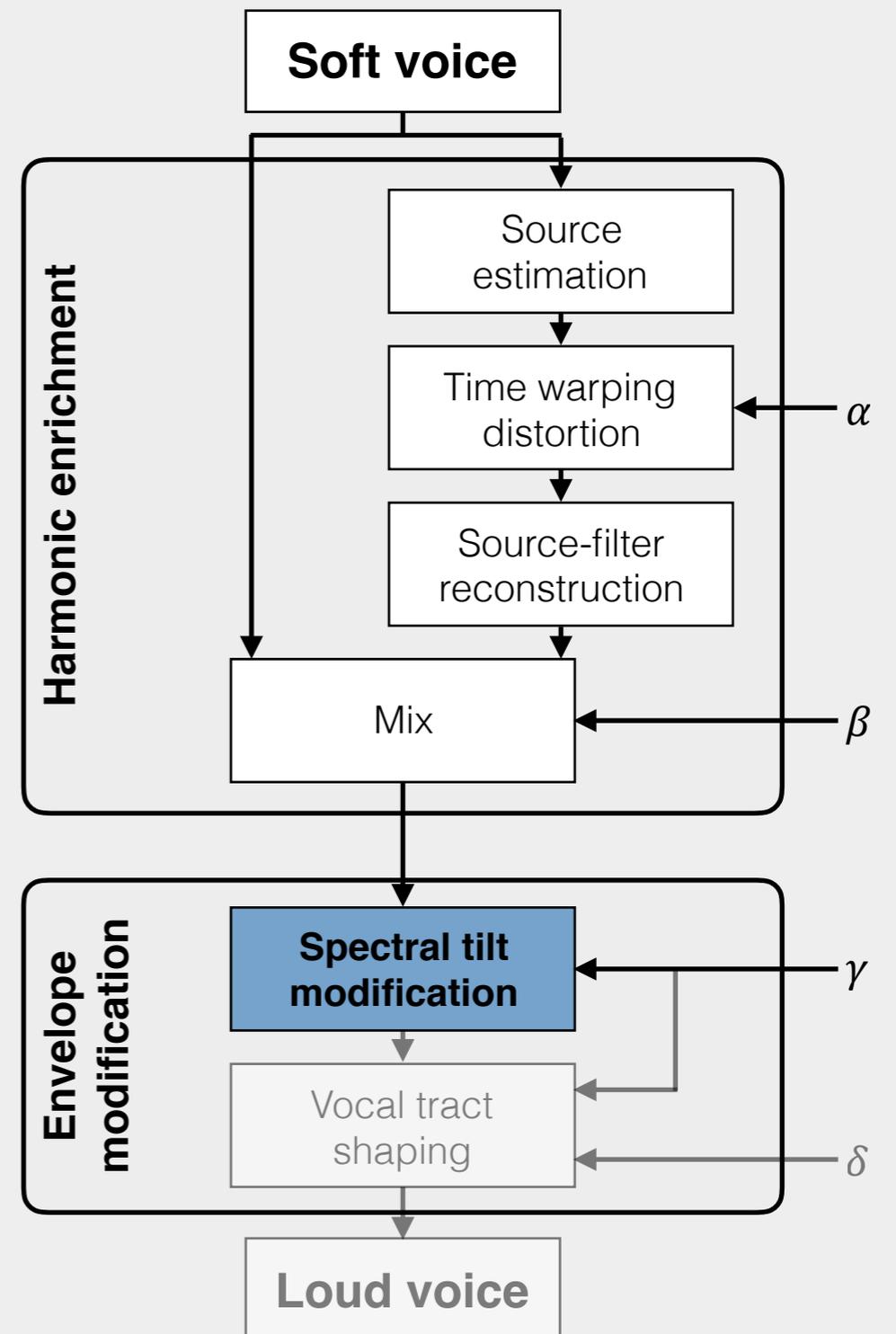
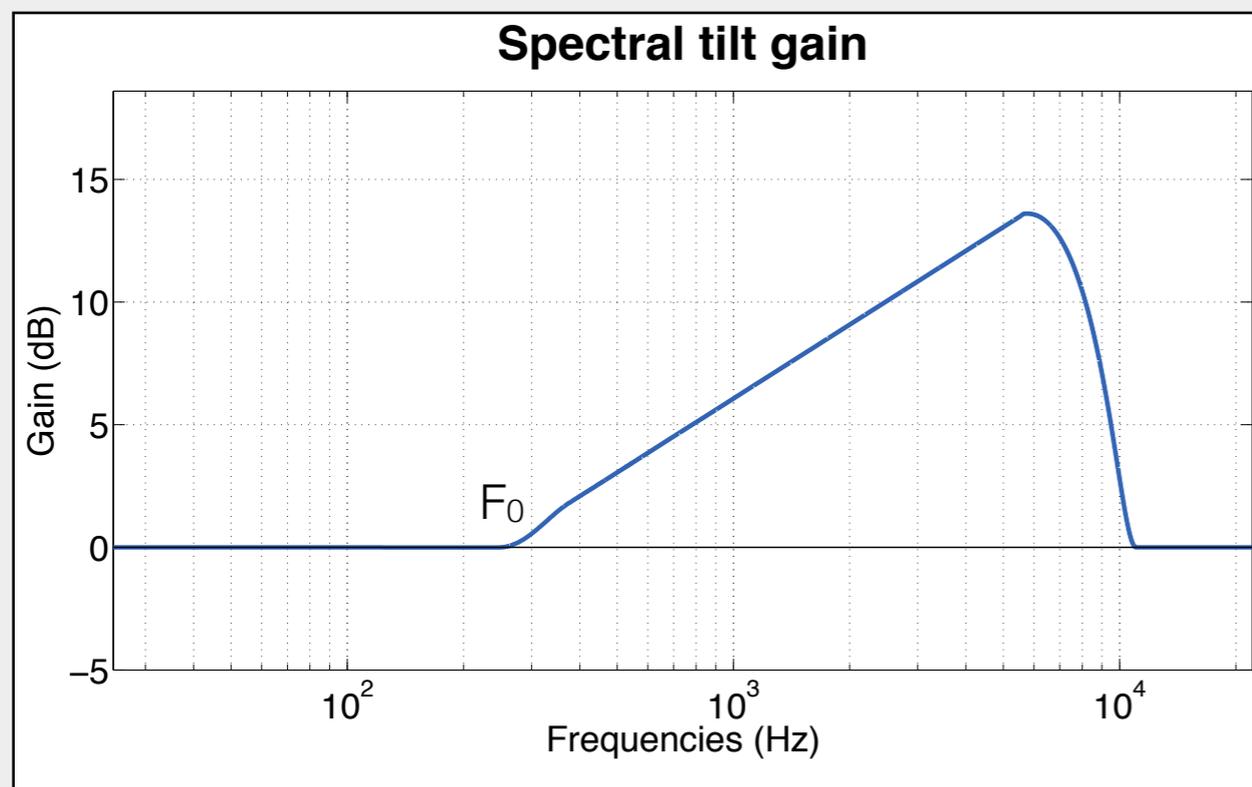
- Introduce new harmonics only where they are missing
 - Design a frequency window on a specific band [Voiced Unvoiced Frequency-10kHz]
 - Mix the original and distorted signals' spectra on this band with a gain β
- New signal with generated harmonic on a specific frequency band



Envelope modification

Spectral tilt

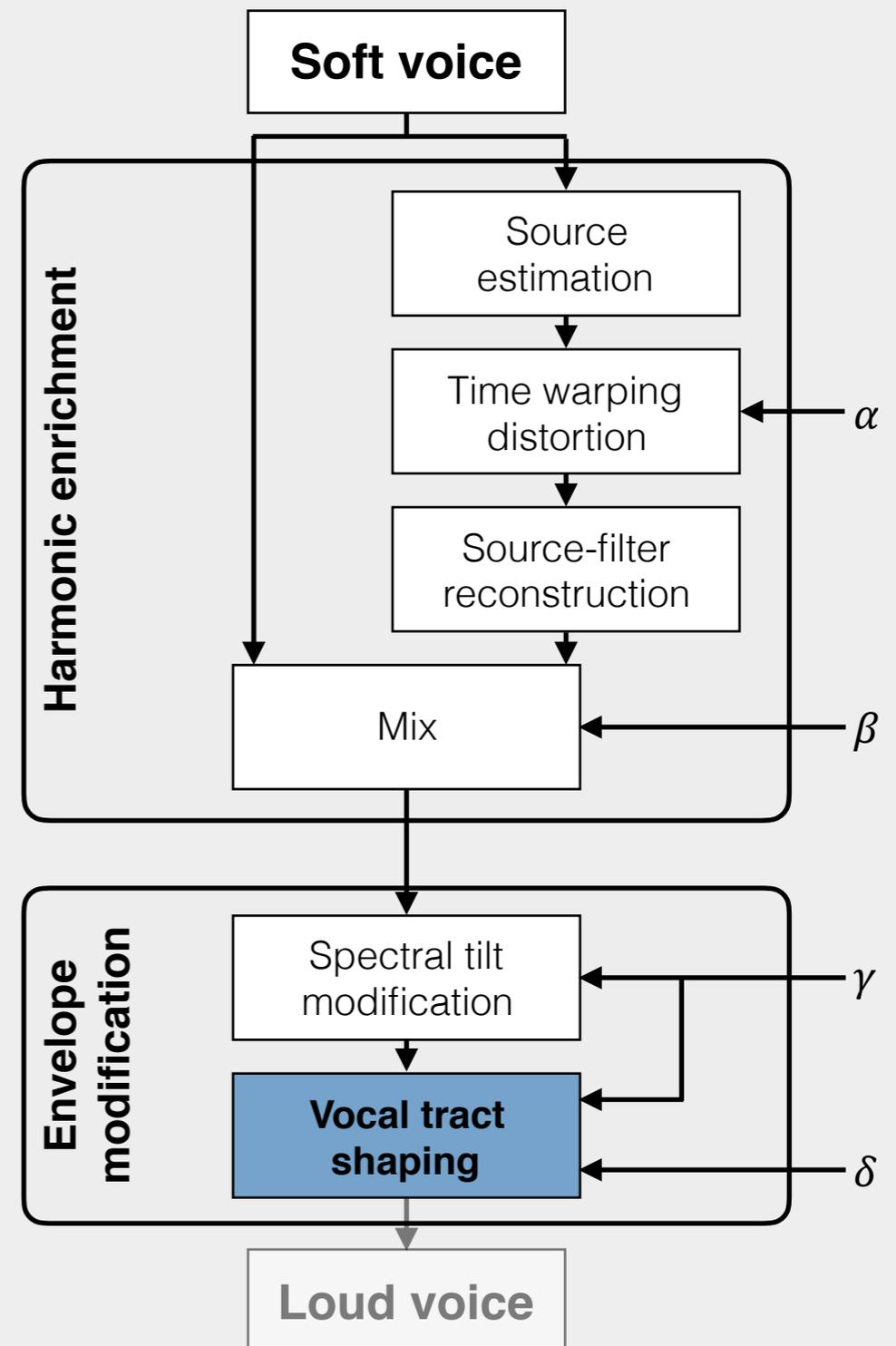
- Louder higher frequencies with vocal effort increase
 - Apply a gain in dB linearly dependent on the log-frequency with a slope γ on $[F_0 - 10\text{kHz}]$



Envelope modification

Formants

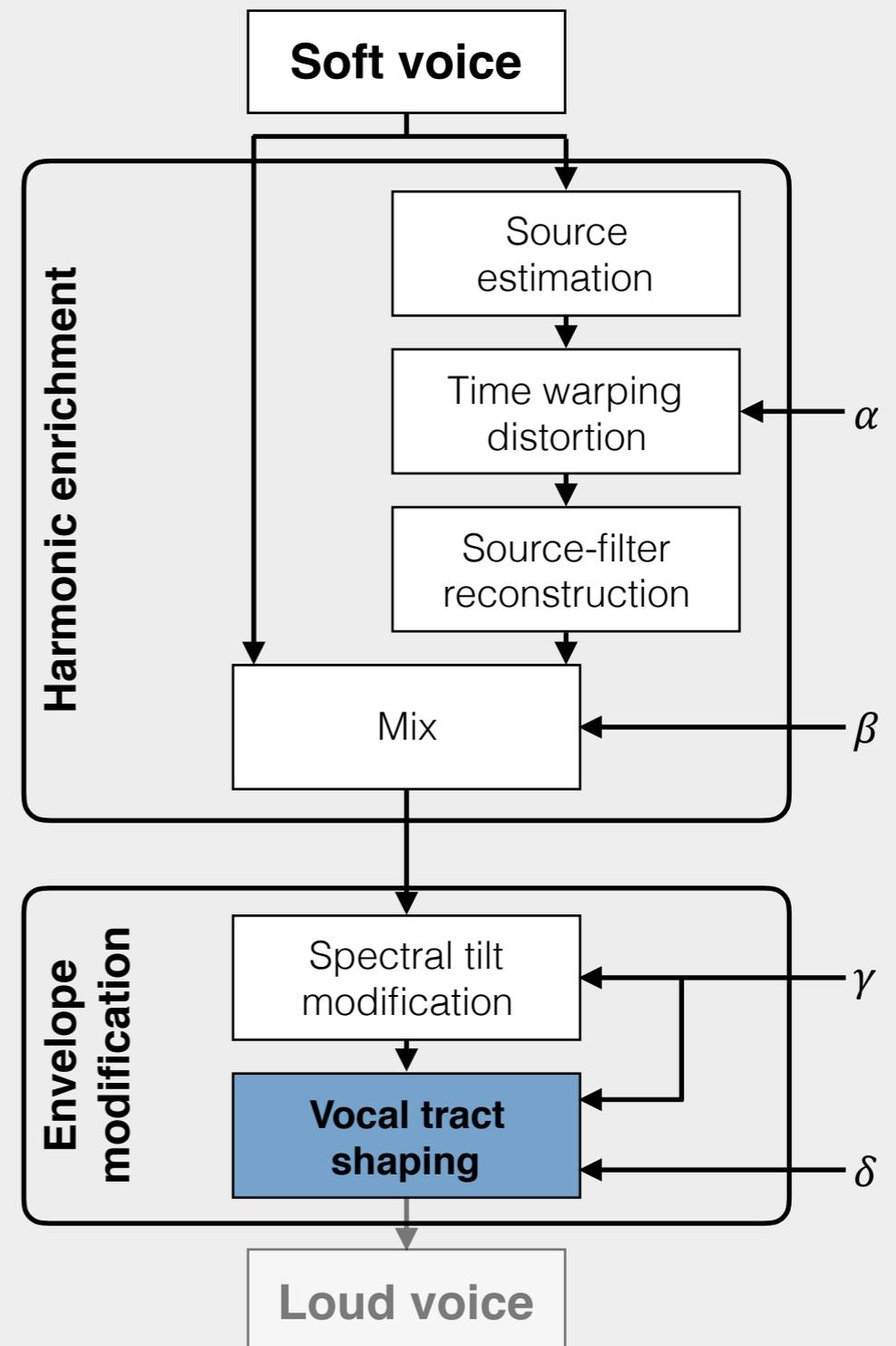
- Better intelligibility with high effort
 - ➔ Extract formant center frequencies, bandwidths, and amplitude from original signal
 - ➔ Increase amplitude of formants with spectral tilt gain γ and additional gain δ if needed
 - ➔ Filter the signal with 5 parallel resonant filters

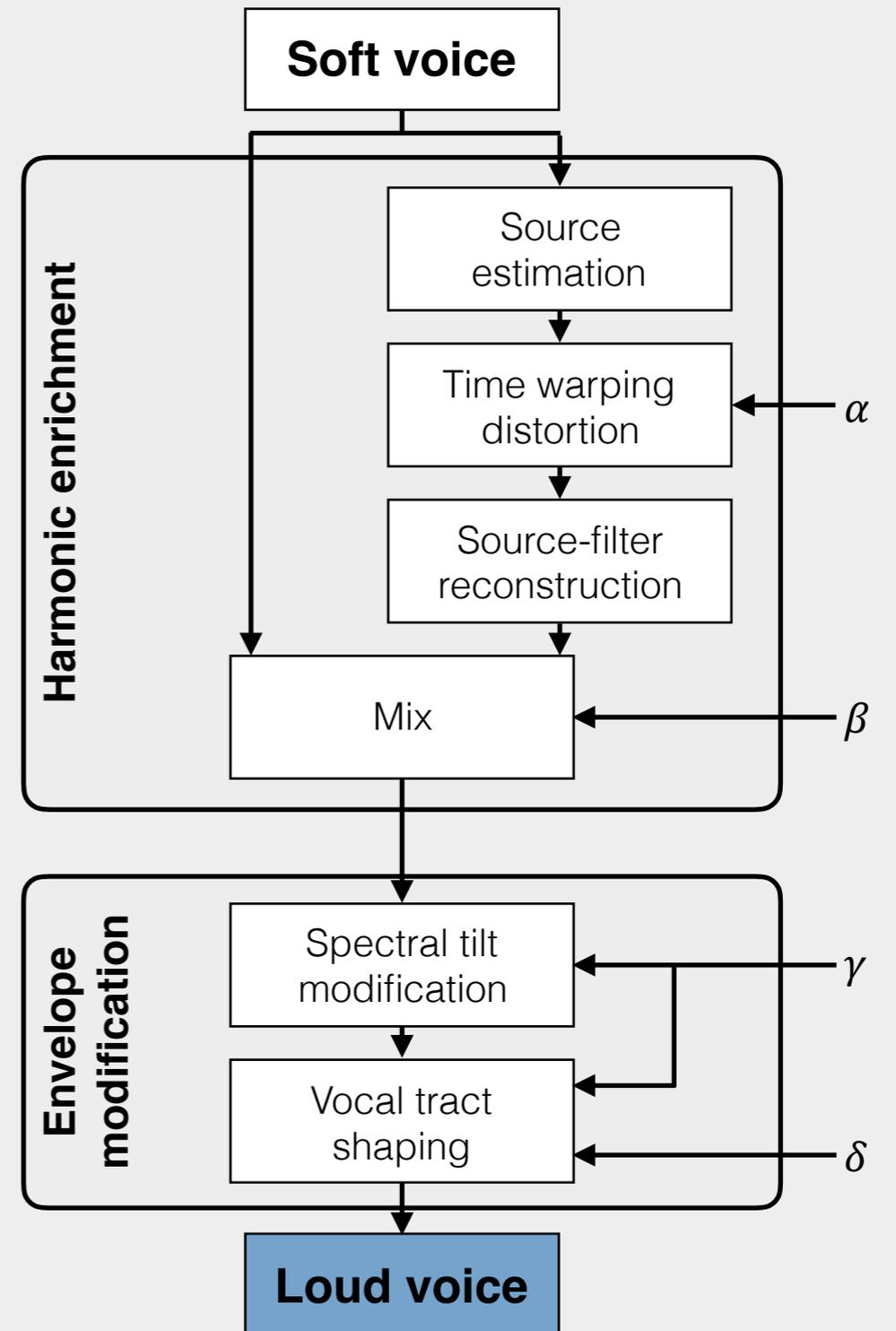


Envelope modification

Formants

- Better intelligibility with high effort
 - ➔ Extract formant center frequencies, bandwidths, and amplitude from original signal
 - ➔ Increase amplitude of formants with spectral tilt gain γ and additional gain δ if needed
 - ➔ Filter the signal with 5 parallel resonant filters
- First formant position naturally increases with vocal effort
 - ➔ Change the position of the first formant by the addition of 10 Hz / dB





Experiment Protocol

Experiment Protocol

- Corpus of isolated vowels
 - Male (baritone) and female (soprano) singers
 - 3 vowels (/a/, /i/, /u/)
 - 3 pitches for the female singer and 1 for the male singer
 - 2 vocal effort levels were selected: pianissimo (to be transformed) and fortissimo (as a reference)

Experiment Protocol

- Corpus of isolated vowels
 - Male (baritone) and female (soprano) singers
 - 3 vowels (/a/, /i/, /u/)
 - 3 pitches for the female singer and 1 for the male singer
 - 2 vocal effort levels were selected: pianissimo (to be transformed) and fortissimo (as a reference)
- 4 conditions to increase the vocal effort of the pianissimo stimuli
 - With spectral modeling
 - With spectral morphing: apply spectral envelope of loud voice on soft voice
 - With / without harmonic enrichment in both methods

Experiment Protocol

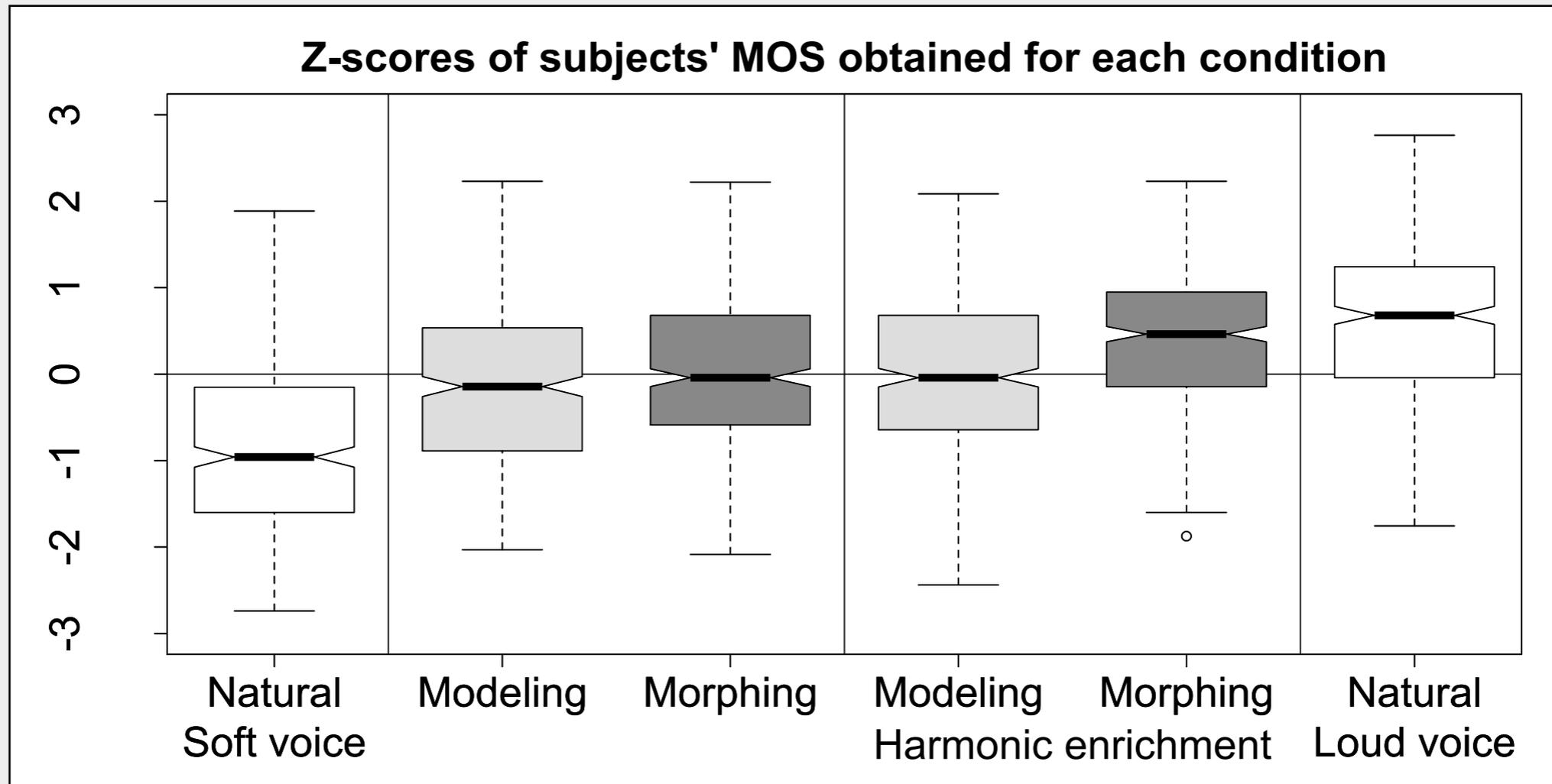
- Corpus of isolated vowels
 - Male (baritone) and female (soprano) singers
 - 3 vowels (/a/, /i/, /u/)
 - 3 pitches for the female singer and 1 for the male singer
 - 2 vocal effort levels were selected: pianissimo (to be transformed) and fortissimo (as a reference)
- 4 conditions to increase the vocal effort of the pianissimo stimuli
 - With spectral modeling
 - With spectral morphing: apply spectral envelope of loud voice on soft voice
 - With / without harmonic enrichment in both methods
- ➔ 90 stimuli (transformed and natural) which were RMS normalized

Experiment Protocol

- Corpus of isolated vowels
 - Male (baritone) and female (soprano) singers
 - 3 vowels (/a/, /i/, /u/)
 - 3 pitches for the female singer and 1 for the male singer
 - 2 vocal effort levels were selected: pianissimo (to be transformed) and fortissimo (as a reference)
- 4 conditions to increase the vocal effort of the pianissimo stimuli
 - With spectral modeling
 - With spectral morphing: apply spectral envelope of loud voice on soft voice
 - With / without harmonic enrichment in both methods
- ➔ 90 stimuli (transformed and natural) which were RMS normalized
- Mean Opinion Score (MOS)
 - 25 subjects
 - Rating the perceived vocal effort of stimuli on a 1 to 5 scale

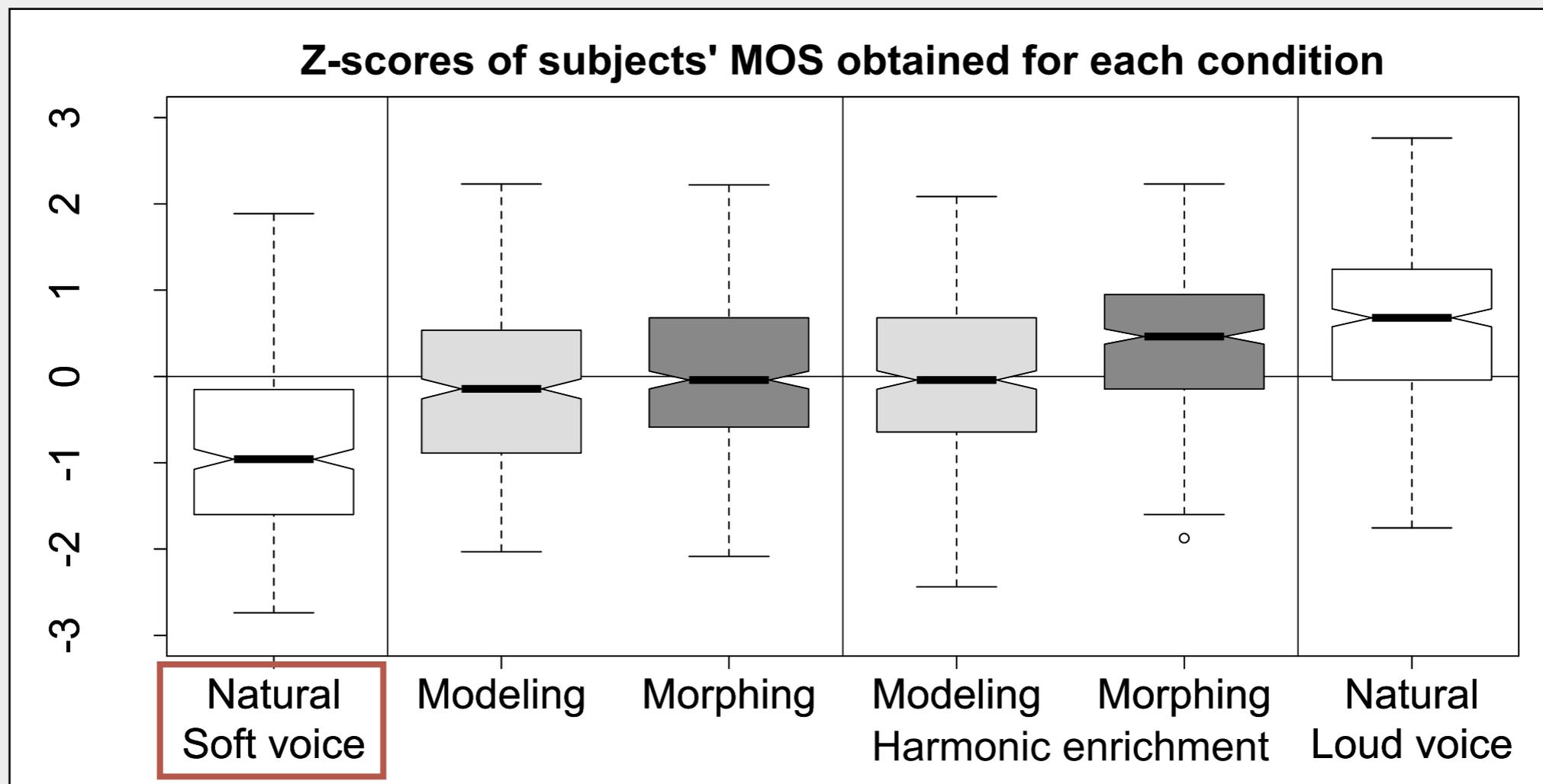
Experiment Results

Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



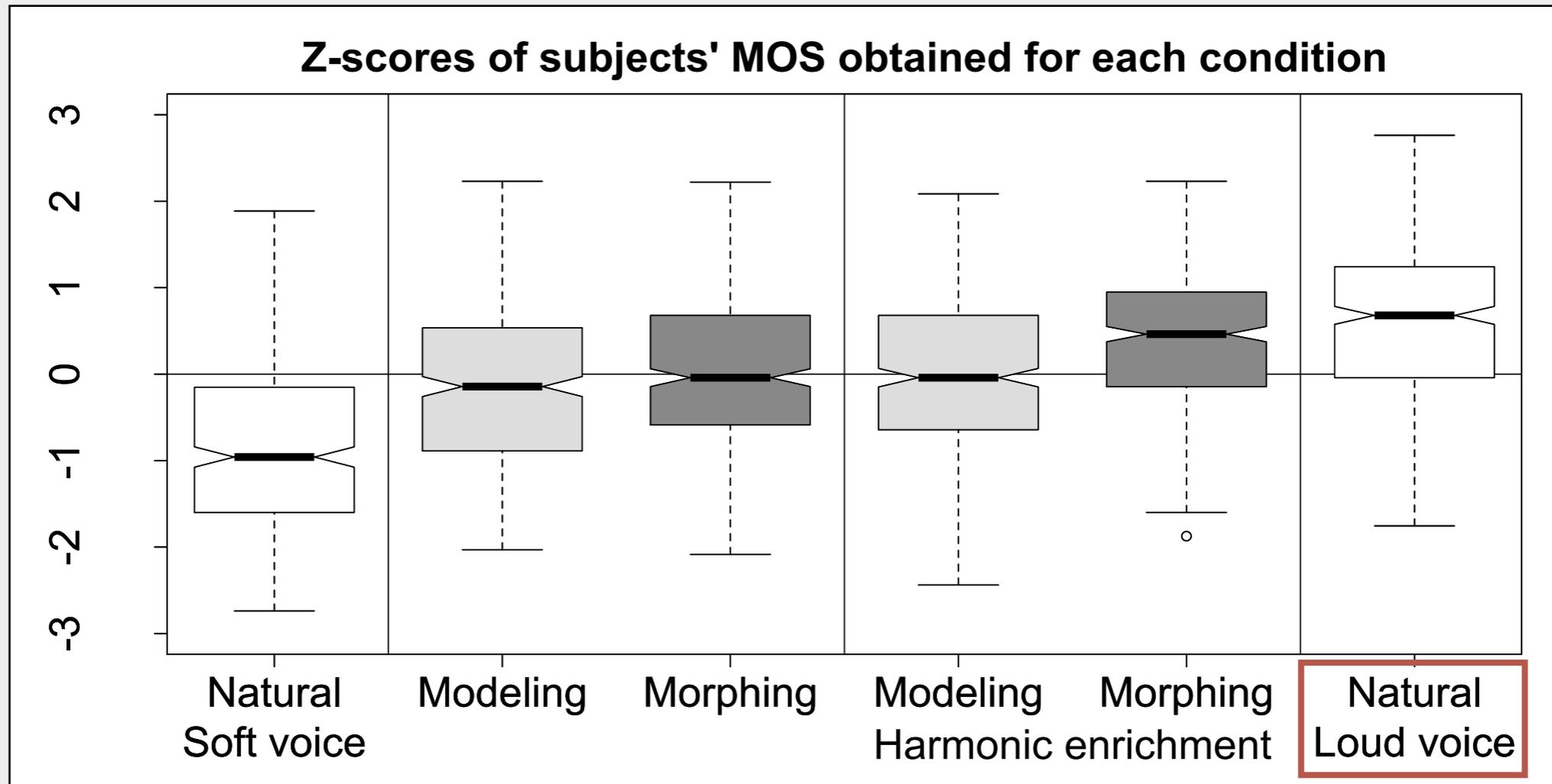
Experiment Results

Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



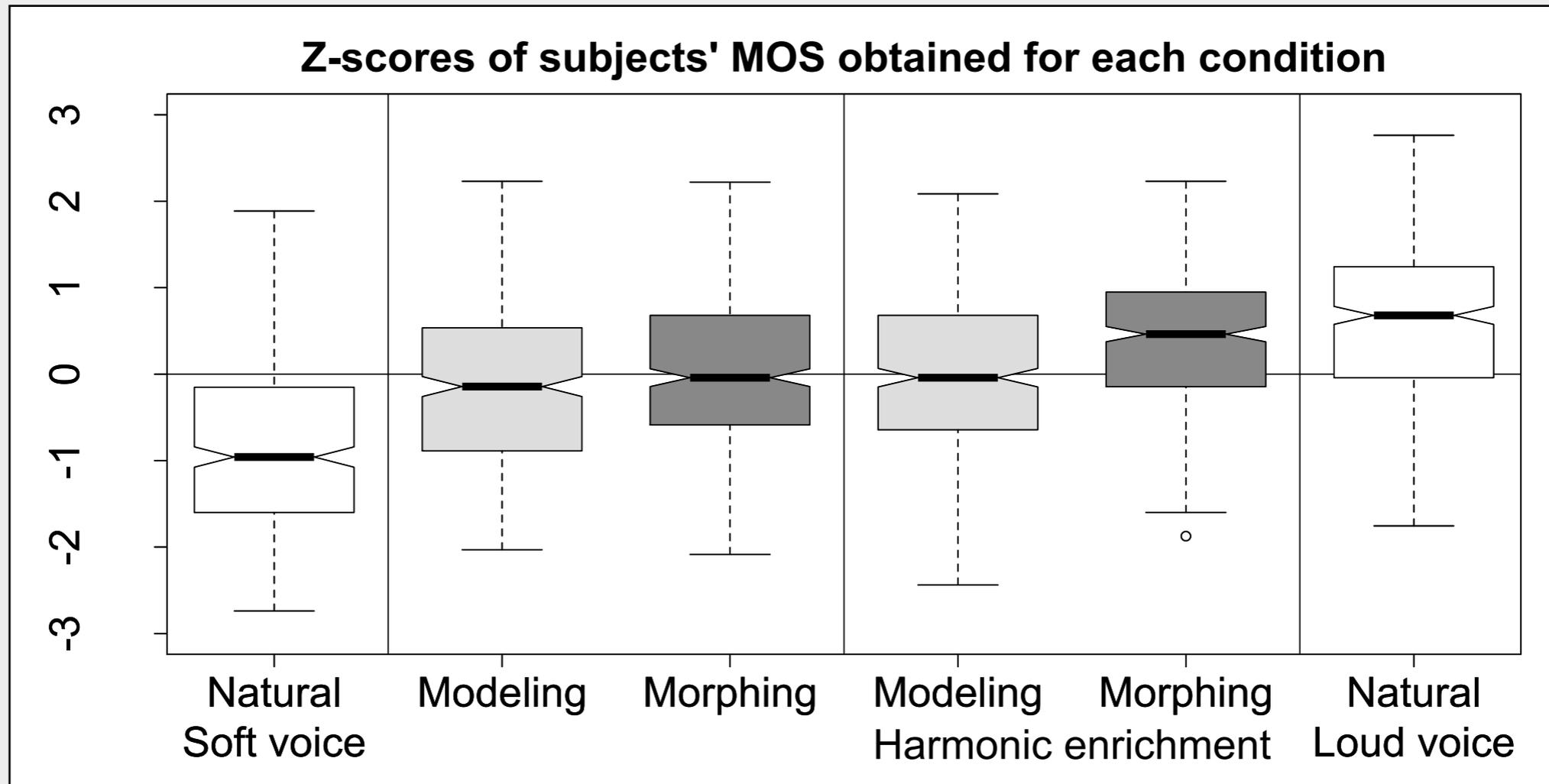
Experiment Results

Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



Experiment Results

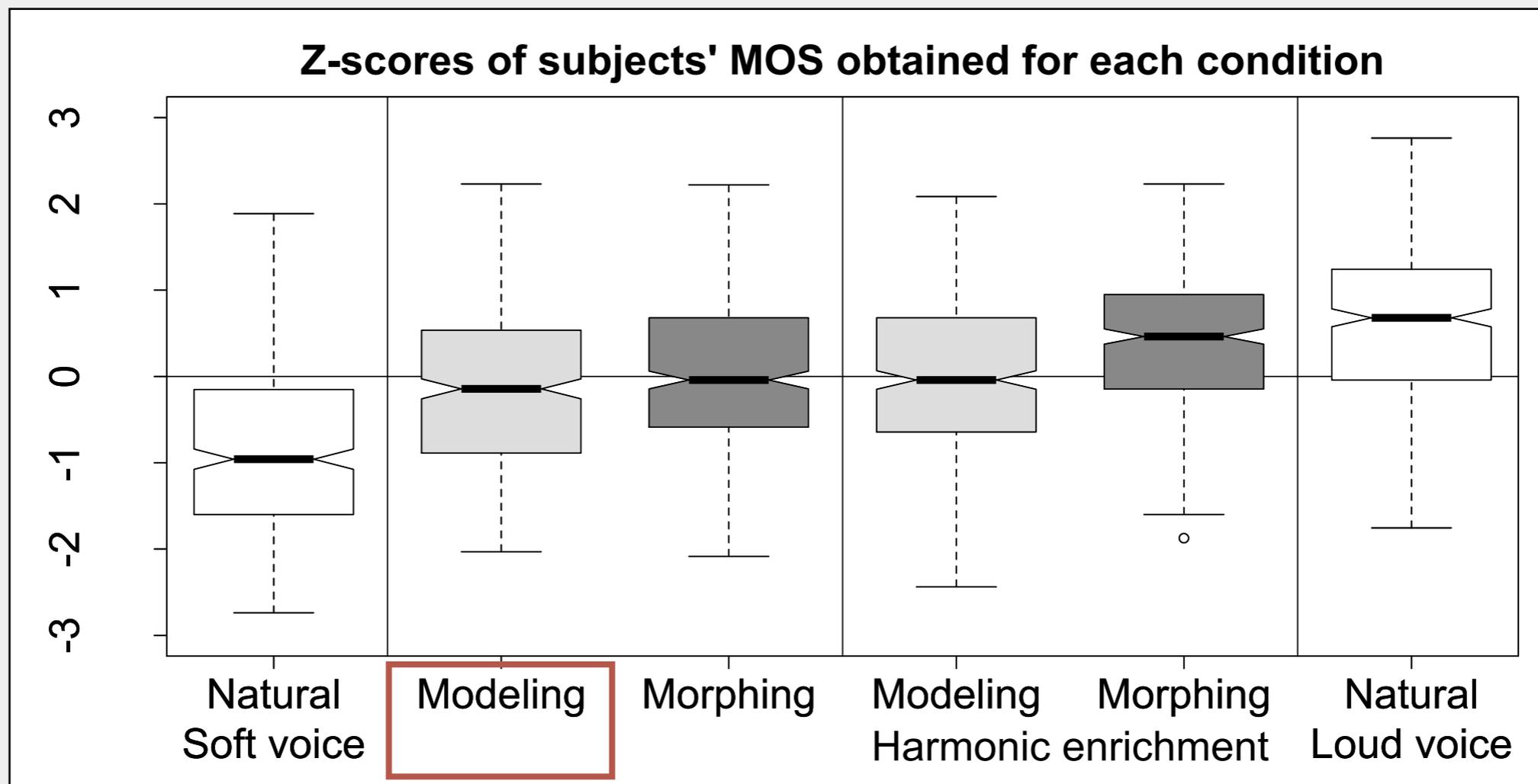
Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



- Morphing method provides the perception of higher vocal effort than modeling

Experiment Results

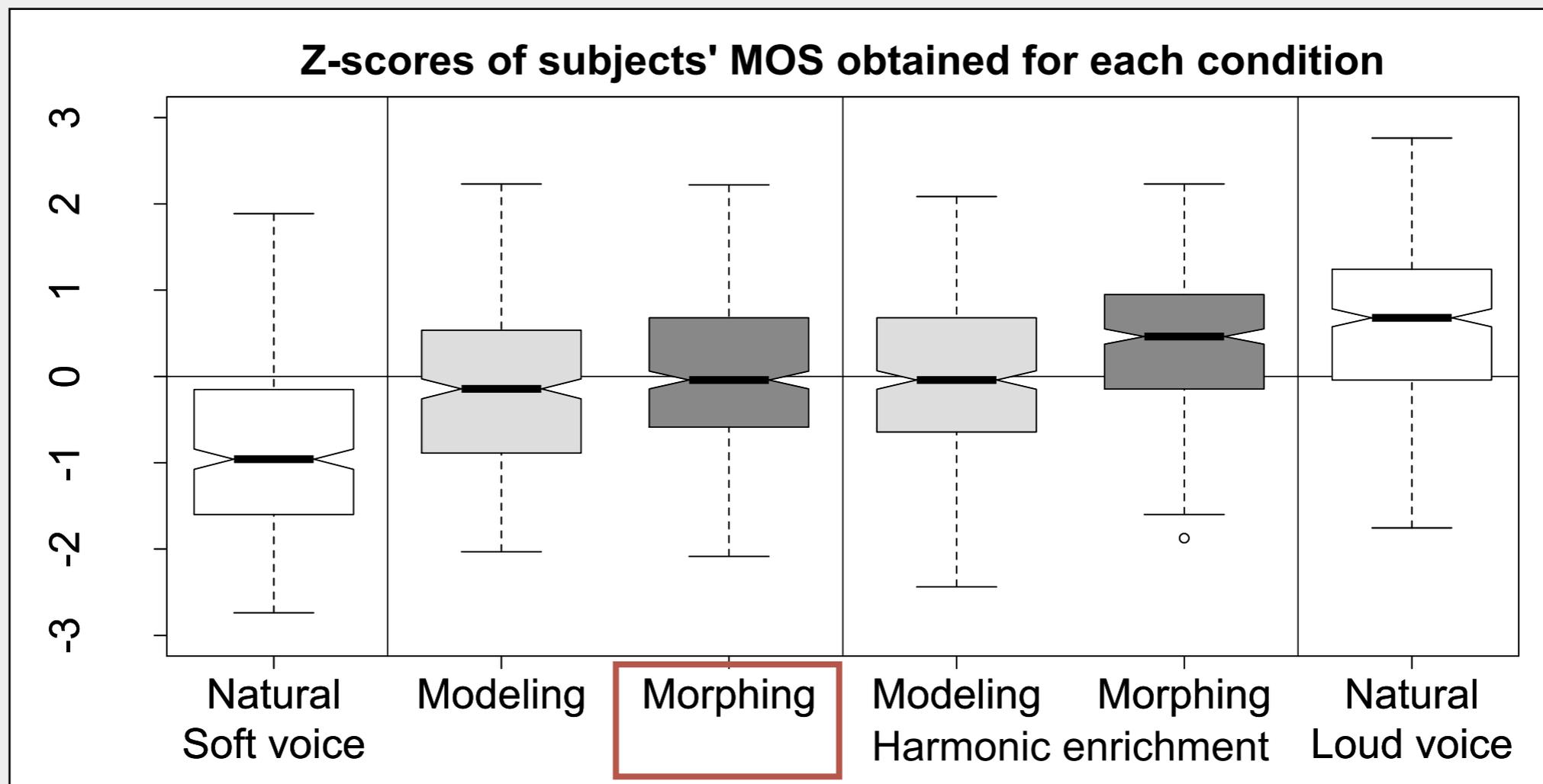
Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



- Morphing method provides the perception of higher vocal effort than modeling

Experiment Results

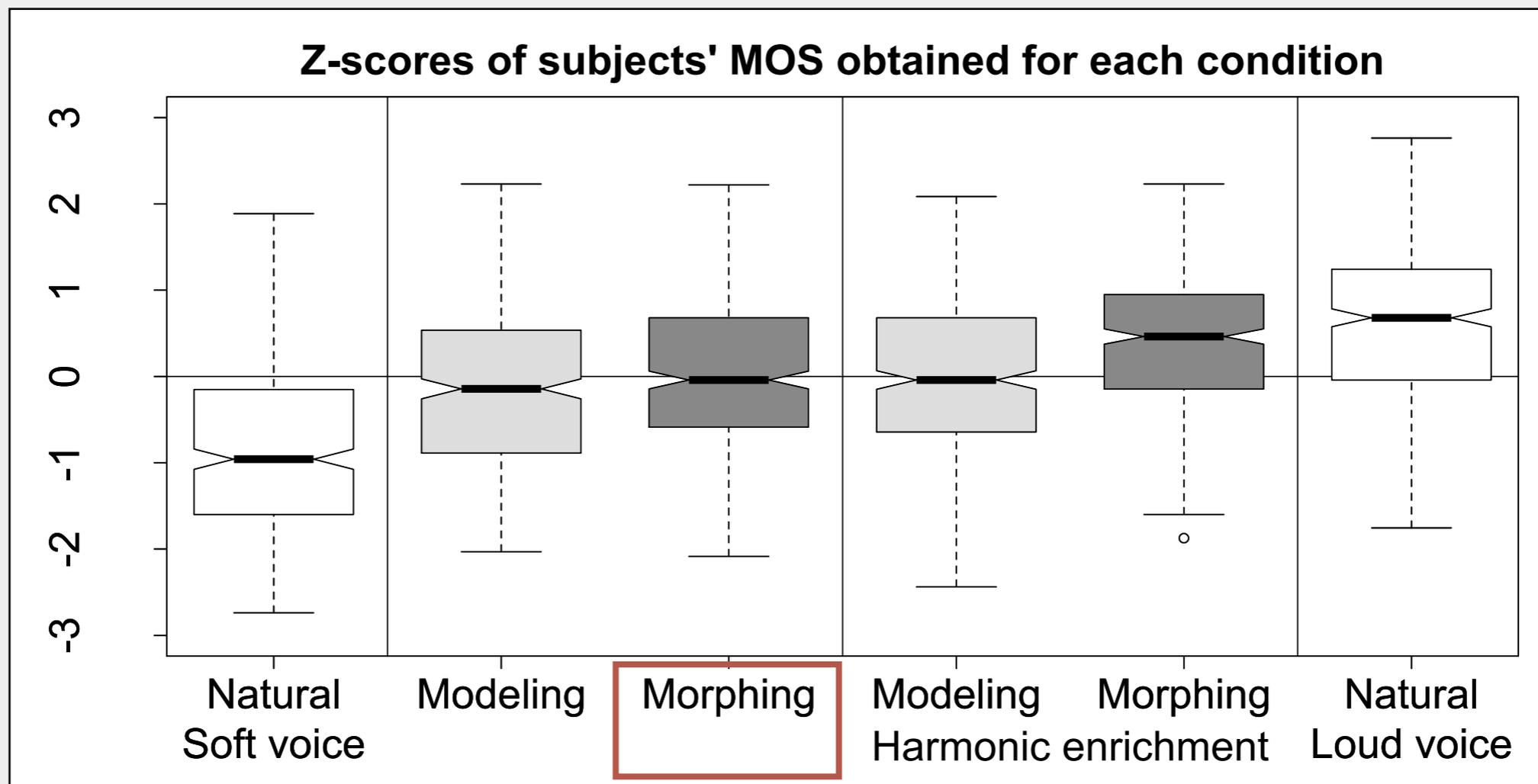
Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



- Morphing method provides the perception of higher vocal effort than modeling

Experiment Results

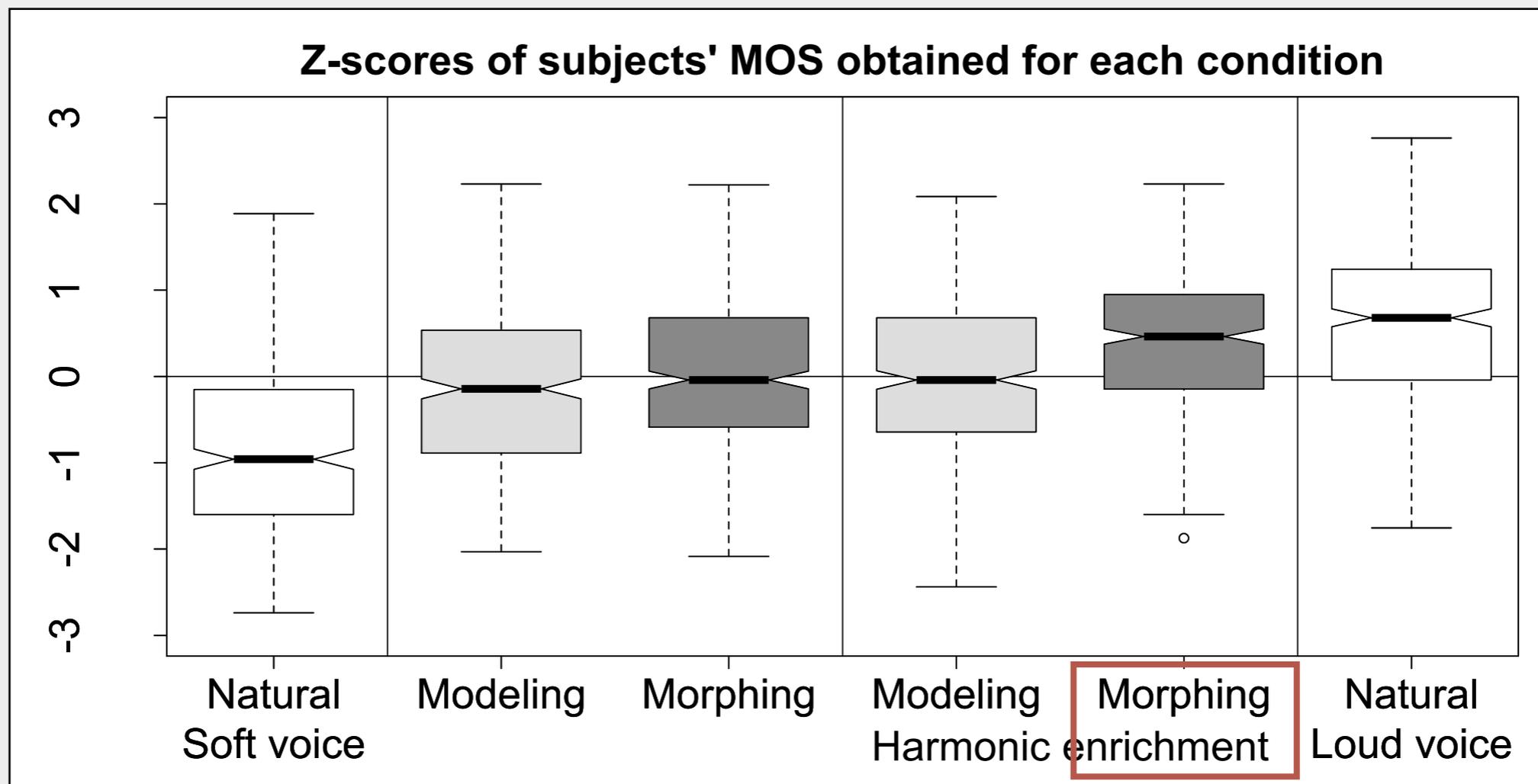
Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



- Morphing method provides the perception of higher vocal effort than modeling
- Harmonic enrichment enables the perception of a higher vocal effort

Experiment Results

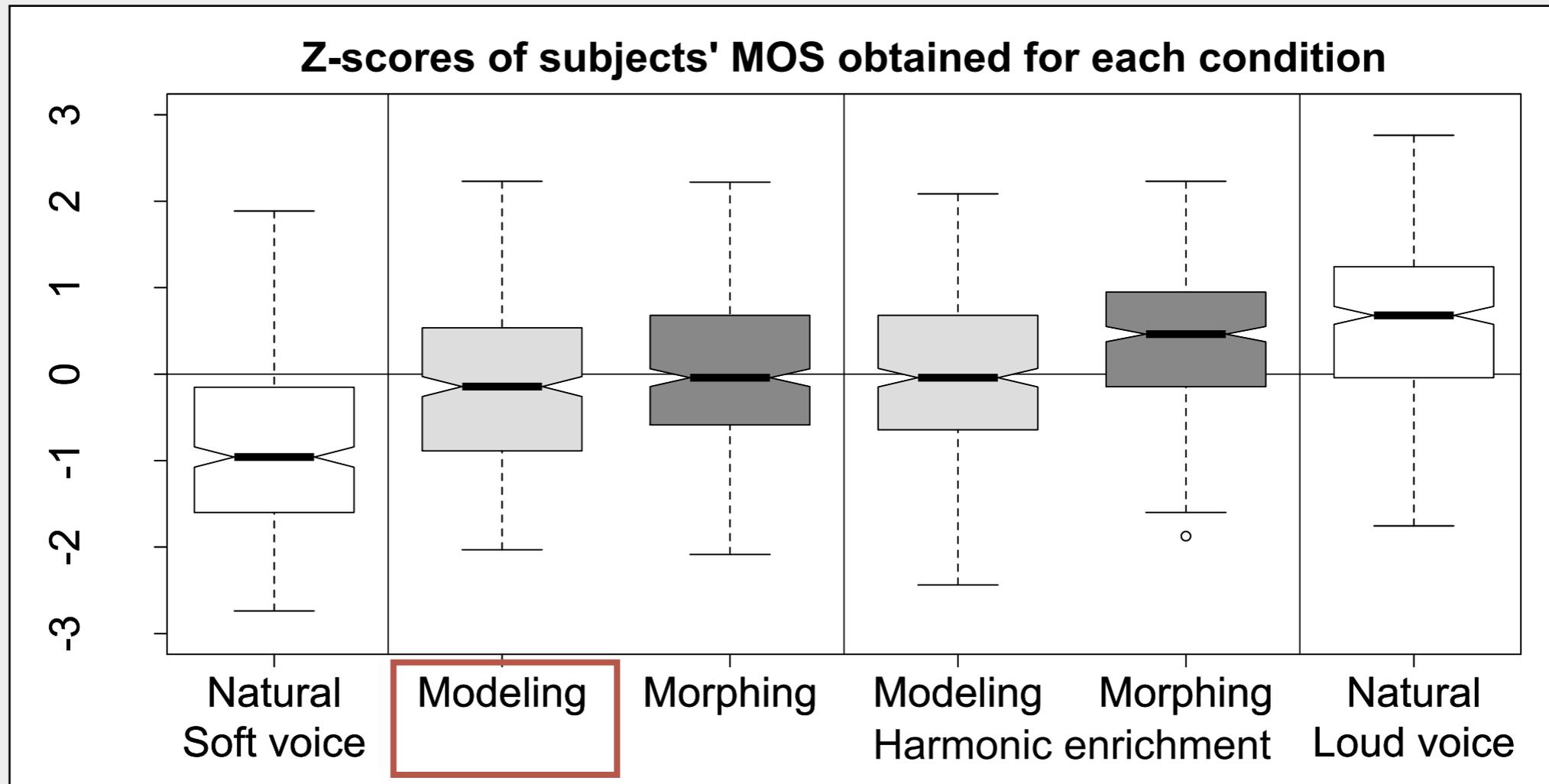
Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



- Morphing method provides the perception of higher vocal effort than modeling
- Harmonic enrichment enables the perception of a higher vocal effort

Experiment Results

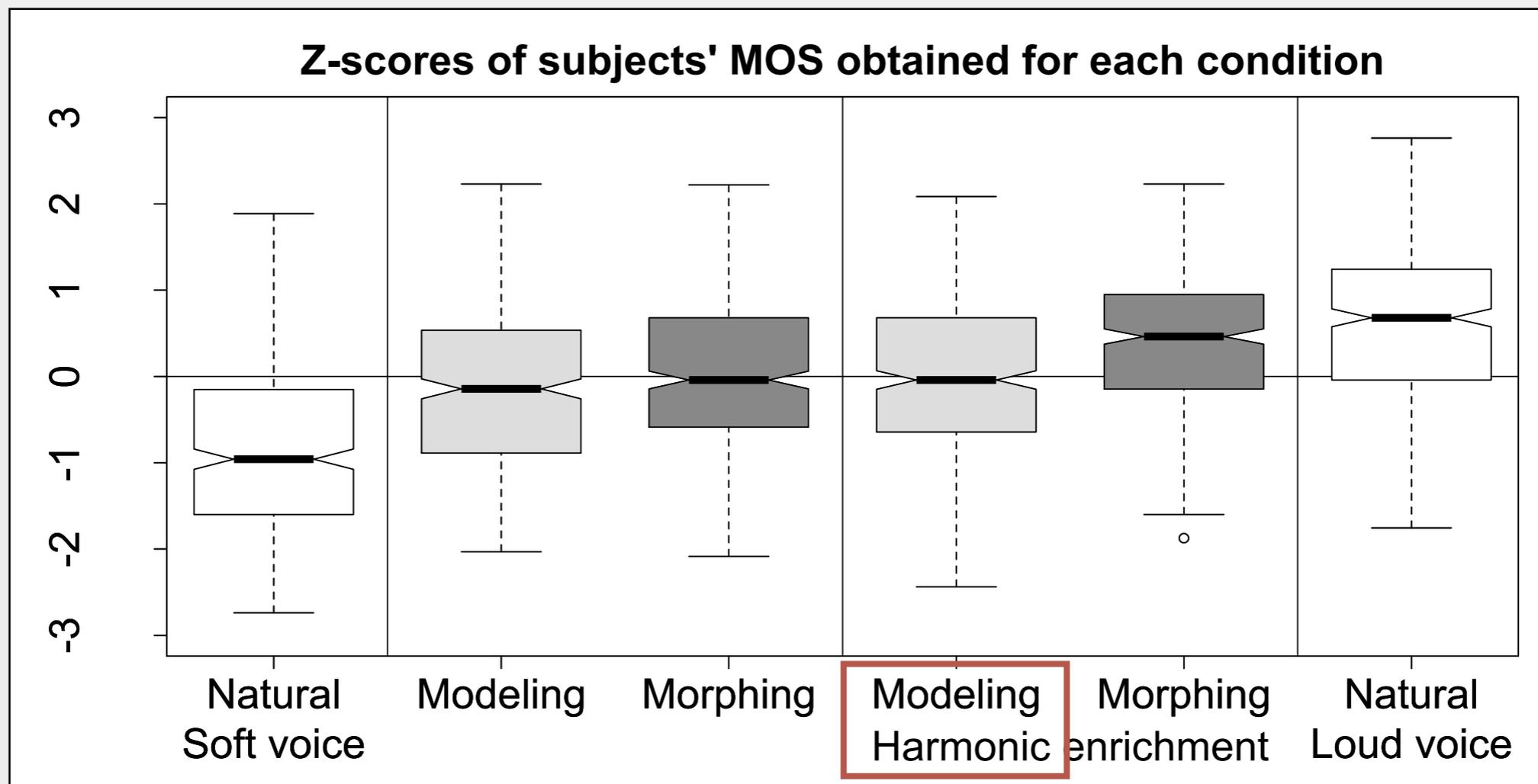
Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



- Morphing method provides the perception of higher vocal effort than modeling
- Harmonic enrichment enables the perception of a higher vocal effort

Experiment Results

Q: Rate the vocal effort of the singer on a scale from 1 to 5 (very low, low, medium, high, very high)



- Morphing method provides the perception of higher vocal effort than modeling
- Harmonic enrichment enables the perception of a higher vocal effort

Conclusion

- Design of a method for vocal effort increase
- Spectral modeling is not as powerful as spectral morphing
 - (-) Our model does not explain all spectral features responsible for vocal effort perception
 - (+) Our model does not need a loud voice target to increase vocal effort
- Harmonic enrichment is essential in the perception of vocal effort

Thank you for your attention !

I'm looking for a postdoc :)