

SPEECH MODIFICATIONS

LECTURE 1: INTRODUCTION

Yannis Stylianou



Telecommunications and Network Lab, ICS- FORTH
styliano@ics.forth.gr

Vitoria, 2010 Sept 2nd

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

1 MODIFICATIONS

2 MODELING AND MODIFICATIONS

- Source modification
- Filter Modification
- Challenges

3 REFERENCES

SPEECH MODIFICATION (OR TRANSFORMATION)

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

DEFINITION (SPEECH MODIFICATION)

Speech Modification aims at the control of non-linguistic information of speech signals such as voice quality and/or voice individuality. By Speech Modification we refer to the various modifications one may apply to the sound produced by a person, speaking or singing.

WHAT IS FOR?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- **Speech Synthesis: past, present, future**
 - Film and Music industry
 - Voice editing/dubbing
 - Toys, Chat rooms, and Games
 - Communications: interpreting telephony, helium speech
 - Voice Pathology: voice restoration
 - High-end hearing aids
 - Other applications include: confuse speaker identification systems.

WHAT IS FOR?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speech Synthesis: past, present, future
- Film and Music industry
 - Voice editing/dubbing
 - Toys, Chat rooms, and Games
 - Communications: interpreting telephony, helium speech
 - Voice Pathology: voice restoration
 - High-end hearing aids
 - Other applications include: confuse speaker identification systems.

WHAT IS FOR?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speech Synthesis: past, present, future
- Film and Music industry
- Voice editing/dubbing
 - Toys, Chat rooms, and Games
 - Communications: interpreting telephony, helium speech
 - Voice Pathology: voice restoration
 - High-end hearing aids
 - Other applications include: confuse speaker identification systems.

WHAT IS FOR?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speech Synthesis: past, present, future
- Film and Music industry
- Voice editing/dubbing
- Toys, Chat rooms, and Games
- Communications: interpreting telephony, helium speech
- Voice Pathology: voice restoration
- High-end hearing aids
- Other applications include: confuse speaker identification systems.

WHAT IS FOR?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speech Synthesis: past, present, future
- Film and Music industry
- Voice editing/dubbing
- Toys, Chat rooms, and Games
- Communications: interpreting telephony, helium speech
 - Voice Pathology: voice restoration
 - High-end hearing aids
 - Other applications include: confuse speaker identification systems.

WHAT IS FOR?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speech Synthesis: past, present, future
- Film and Music industry
- Voice editing/dubbing
- Toys, Chat rooms, and Games
- Communications: interpreting telephony, helium speech
- Voice Pathology: voice restoration
- High-end hearing aids
- Other applications include: confuse speaker identification systems.

WHAT IS FOR?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speech Synthesis: past, present, future
- Film and Music industry
- Voice editing/dubbing
- Toys, Chat rooms, and Games
- Communications: interpreting telephony, helium speech
- Voice Pathology: voice restoration
- High-end hearing aids
- Other applications include: confuse speaker identification systems.

WHAT IS FOR?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speech Synthesis: past, present, future
- Film and Music industry
- Voice editing/dubbing
- Toys, Chat rooms, and Games
- Communications: interpreting telephony, helium speech
- Voice Pathology: voice restoration
- High-end hearing aids
- Other applications include: confuse speaker identification systems.

COVERED AREAS OF RESEARCH

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- *From* speech production modeling and understanding
To perception of speech
- *From* natural language processing, modeling and control
of speaking style
To pattern recognition and statistical signal processing.

CONNECTIONS WITH OTHER SPEECH AREAS

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speaker identification
- Speech modeling/coding
- Speech recognition
- Speech enhancement
- Speech Synthesis

CONNECTIONS WITH OTHER SPEECH AREAS

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speaker identification
- Speech modeling/coding
 - Speech recognition
 - Speech enhancement
 - Speech Synthesis

CONNECTIONS WITH OTHER SPEECH AREAS

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speaker identification
- Speech modeling/coding
- Speech recognition
- Speech enhancement
- Speech Synthesis

CONNECTIONS WITH OTHER SPEECH AREAS

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speaker identification
- Speech modeling/coding
- Speech recognition
- Speech enhancement
- Speech Synthesis

CONNECTIONS WITH OTHER SPEECH AREAS

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Speaker identification
- Speech modeling/coding
- Speech recognition
- Speech enhancement
- Speech Synthesis

WAYS TO MODIFY/TRANSFORM A VOICE (OR SPEECH)

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Time-scale modification
- Pitch modification
- Speaker modification/Voice Alteration
- Voice quality control
- Voice morphing
- Voice conversion

WAYS TO MODIFY/TRANSFORM A VOICE (OR SPEECH)

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Time-scale modification
- **Pitch modification**
 - Speaker modification/Voice Alteration
 - Voice quality control
 - Voice morphing
 - Voice conversion

WAYS TO MODIFY/TRANSFORM A VOICE (OR SPEECH)

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Time-scale modification
- Pitch modification
- **Speaker modification/Voice Alteration**
 - Voice quality control
 - Voice morphing
 - Voice conversion

WAYS TO MODIFY/TRANSFORM A VOICE (OR SPEECH)

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Time-scale modification
- Pitch modification
- Speaker modification/Voice Alteration
- **Voice quality control**
- Voice morphing
- Voice conversion

WAYS TO MODIFY/TRANSFORM A VOICE (OR SPEECH)

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Time-scale modification
- Pitch modification
- Speaker modification/Voice Alteration
- Voice quality control
- **Voice morphing**
- Voice conversion

WAYS TO MODIFY/TRANSFORM A VOICE (OR SPEECH)

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Time-scale modification
- Pitch modification
- Speaker modification/Voice Alteration
- Voice quality control
- Voice morphing
- **Voice conversion**

VOICE INDIVIDUALITY

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Socio/psychological dimension: speaking style, emotional state
- Physiological dimension: voice quality

VOICE INDIVIDUALITY

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

- Socio/psychological dimension: speaking style, emotional state
- Physiological dimension: voice quality

SPEAKING STYLE VERSUS VOICE QUALITY

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Speaking style is mainly acoustically realized in prosodic features such as:

- pitch contour
- duration of words
- rhythm, pauses
- power levels

Voice quality is mostly reflected in the power spectrum of the:

- glottal source signal
- vocal tract filter (including nasal cavity and mouth)

SPEAKING STYLE VERSUS VOICE QUALITY

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Speaking style is mainly acoustically realized in prosodic features such as:

- pitch contour
- duration of words
- rhythm, pauses
- power levels

Voice quality is mostly reflected in the power spectrum of the:

- glottal source signal
- vocal tract filter (including nasal cavity and mouth)

ACOUSTIC CHARACTERISTICS OF VOICE INDIVIDUALITY

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Voice source:

- glottal wave shape
- average fundamental frequency
- fundamental frequency contour (pitch contour)
- fundamental frequency fluctuations

Vocal tract filter:

- average speech spectrum
- shape of spectral envelope and spectral tilt
- formant frequencies, bandwidths
- formant trajectories

ACOUSTIC CHARACTERISTICS OF VOICE INDIVIDUALITY

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Voice source:

- glottal wave shape
- average fundamental frequency
- fundamental frequency contour (pitch contour)
- fundamental frequency fluctuations

Vocal tract filter:

- average speech spectrum
- shape of spectral envelope and spectral tilt
- formant frequencies, bandwidths
- formant trajectories

ANY GOLD CHARACTERISTIC FOR VOICE INDIVIDUALITY?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Defining Voice Individuality [1]:

- **Matsumoto et al. 1973: fundamental freq., formant freq., spectral tilt**
- Sato 1974 and Karlsson 1986, 1991: average speech spectrum
- Furui, 1986: average spectrum, fundamental freq.
- Nakatsui, 1976: fundamental freq., formant freq.
- Itoh, 1982: Spectral envelope, fundamental freq.

Summary: A single gold characteristic doesn't exist; individuality is an amalgam of many characteristics.

ANY GOLD CHARACTERISTIC FOR VOICE INDIVIDUALITY?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Defining Voice Individuality [1]:

- Matsumoto et al. 1973: fundamental freq., formant freq., spectral tilt
- Sato 1974 and Karlsson 1986, 1991: average speech spectrum
 - Furui, 1986: average spectrum, fundamental freq.
 - Nakatsui, 1976: fundamental freq., formant freq.
 - Itoh, 1982: Spectral envelope, fundamental freq.

Summary: A single gold characteristic doesn't exist; individuality is an amalgam of many characteristics.

ANY GOLD CHARACTERISTIC FOR VOICE INDIVIDUALITY?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Defining Voice Individuality [1]:

- Matsumoto et al. 1973: fundamental freq., formant freq., spectral tilt
- Sato 1974 and Karlsson 1986, 1991: average speech spectrum
- **Furui, 1986: average spectrum, fundamental freq.**
 - Nakatsui, 1976: fundamental freq., formant freq.
 - Itoh, 1982: Spectral envelope, fundamental freq.

Summary: A single gold characteristic doesn't exist; individuality is an amalgam of many characteristics.

ANY GOLD CHARACTERISTIC FOR VOICE INDIVIDUALITY?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Defining Voice Individuality [1]:

- Matsumoto et al. 1973: fundamental freq., formant freq., spectral tilt
- Sato 1974 and Karlsson 1986, 1991: average speech spectrum
- Furui, 1986: average spectrum, fundamental freq.
- Nakatsui, 1976: fundamental freq., formant freq.
- Itoh, 1982: Spectral envelope, fundamental freq.

Summary: A single gold characteristic doesn't exist; individuality is an amalgam of many characteristics.

ANY GOLD CHARACTERISTIC FOR VOICE INDIVIDUALITY?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Defining Voice Individuality [1]:

- Matsumoto et al. 1973: fundamental freq., formant freq., spectral tilt
- Sato 1974 and Karlsson 1986, 1991: average speech spectrum
- Furui, 1986: average spectrum, fundamental freq.
- Nakatsui, 1976: fundamental freq., formant freq.
- Itoh, 1982: Spectral envelope, fundamental freq.

Summary: A single gold characteristic doesn't exist; individuality is an amalgam of many characteristics.

ANY GOLD CHARACTERISTIC FOR VOICE INDIVIDUALITY?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Defining Voice Individuality [1]:

- Matsumoto et al. 1973: fundamental freq., formant freq., spectral tilt
- Sato 1974 and Karlsson 1986, 1991: average speech spectrum
- Furui, 1986: average spectrum, fundamental freq.
- Nakatsui, 1976: fundamental freq., formant freq.
- Itoh, 1982: Spectral envelope, fundamental freq.

Summary: A single gold characteristic doesn't exist; individuality is an amalgam of many characteristics.

ANY GOLD CHARACTERISTIC FOR VOICE INDIVIDUALITY?

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References

Defining Voice Individuality [1]:

- Matsumoto et al. 1973: fundamental freq., formant freq., spectral tilt
- Sato 1974 and Karlsson 1986, 1991: average speech spectrum
- Furui, 1986: average spectrum, fundamental freq.
- Nakatsui, 1976: fundamental freq., formant freq.
- Itoh, 1982: Spectral envelope, fundamental freq.

Summary: A single gold characteristic doesn't exist; individuality is an amalgam of many characteristics.

MODIFICATIONS AND SOURCE/FILTER THEORY

Speech
Modifications

Yannis
Stylianou

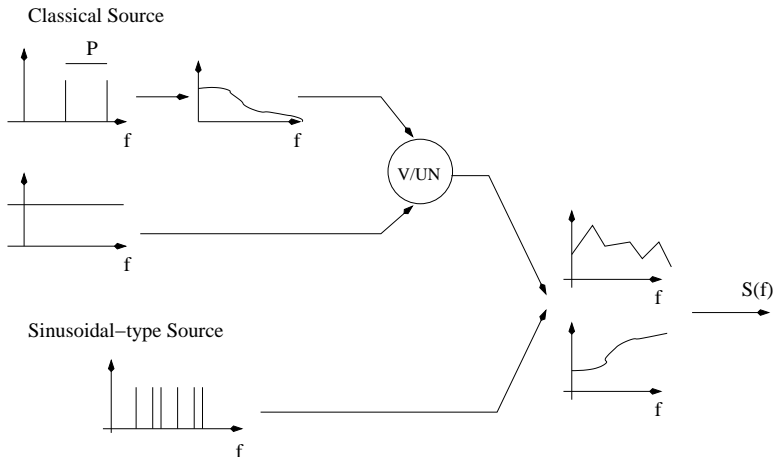
Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification
Filter
Modification
Challenges

References



TIME-SCALE AND PITCH MODIFICATION

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification

Filter
Modification
Challenges

References

DEFINITION (TIME-SCALE MODIFICATION)

In time-scale modification the rate of articulation is changed without affecting the perceptual quality of the original speech.

DEFINITION (PITCH MODIFICATION)

The fundamental frequency is changed while preserving the short-time envelope characteristics as well as the duration of the original speech.

TIME-SCALE AND PITCH MODIFICATION

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification

Filter
Modification
Challenges

References

DEFINITION (TIME-SCALE MODIFICATION)

In time-scale modification the rate of articulation is changed without affecting the perceptual quality of the original speech.

DEFINITION (PITCH MODIFICATION)

The fundamental frequency is changed while preserving the short-time envelope characteristics as well as the duration of the original speech.

PROPOSED SYSTEMS FOR TIME-SCALE MODIFICATION

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification

Filter
Modification
Challenges

References

- TD-PSOLA [2],
- WSOLA [3],
- HNM [4],
- STC [5],
- STRAIGHT [6]

PROPOSED SYSTEMS FOR PITCH MODIFICATION

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification

Filter
Modification
Challenges

References

- TD-PSOLA [2],
- HNM [4],
- STC [5],
- STRAIGHT [6]

FILTER MODIFICATION OVERVIEW

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification

Filter
Modification
Challenges

References

- Abe et al. (1988)[7]: VQ mapping
- Valbret et al. (1992)[8]: Linear Multivariate Regression (LMR), Dynamic Frequency Warping (DFW)
- Iwahashi et al. (1994)[9]: Speaker Interpolation
- Kuwabara et al. (1995)[1]: Fuzzy VQ
- Stylianou et al. (1995)[10]: Probabilistic approach (GMM)
- Kain et al. (1998) [11]: Probabilistic approach (GMM)
- Toda et al. (2001) [12]: Probabilistic approach (GMM)and DFW
- Toda et al. (2005) [13]: Probabilistic approach (GMM)
- Turk et al. (2005) [14]: Correction filters
- Sündermann et al. (2006) [15]: Unit selection
- Mouchtaris et al.(2006)[16]: Probabilistic approach (GMM)and speaker adaptation
- Erro (2008)[17]: Intra and Cross-lingual Voice Conversion based on Harmonic plus Stochastic modeling.

CHALLENGES

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification
Filter
Modification
Challenges

References

- **Quality**
 - Model interactions
 - The cafeteria noise case
 - Changing speaking rate
 - Model speaking style
 - Detection of Transformation

CHALLENGES

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification
Filter
Modification
Challenges

References

- Quality
- **Model interactions**
 - The cafeteria noise case
 - Changing speaking rate
- Model speaking style
- Detection of Transformation

CHALLENGES

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification
Filter
Modification
Challenges

References

- Quality
- Model interactions
 - The cafeteria noise case
 - Changing speaking rate
 - Model speaking style
 - Detection of Transformation

CHALLENGES

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification
Filter
Modification
Challenges

References

- Quality
- Model interactions
 - The cafeteria noise case
 - Changing speaking rate
- Model speaking style
- Detection of Transformation

CHALLENGES

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification
Filter
Modification
Challenges

References

- Quality
- Model interactions
 - The cafeteria noise case
 - Changing speaking rate
- **Model speaking style**
- Detection of Transformation

CHALLENGES

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

Source
modification
Filter
Modification
Challenges

References

- Quality
- Model interactions
 - The cafeteria noise case
 - Changing speaking rate
- Model speaking style
- Detection of Transformation

Speech Modifications

Yannis Stylianou

Outline of the talk

Modifications

Modeling and Modifications

References



H. Kuwabara and Y. Sagisaka, "Acoustic characteristics of speaker individuality: Control and conversion," *Speech Communication*, vol. 16, no. 2, pp. 165–173, 1995.



E. Moulines and J. Laroche, "Techniques for pitch-scale and time-scale transformation of speech. part I. non parametric methods," *Speech Communication*, vol. 16, Feb. 1995.



W. Verhelst and M. Roelands, "An overlap-add technique based on waveform similarity (wsola) for high quality time-scale modification of speech," in *Proc. ICASSP93*, pp. 554–557, 1993.



Y. Stylianou, J. Laroche, and E. Moulines, "High-Quality Speech Modification based on a Harmonic + Noise Model.," *Proc. EUROSPEECH*, 1995.



R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 744–754, Aug 1986.



H. Kuwahara, "Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Munich, Germany), pp. 1303–1306, 1997.



M. Abe, S. Nakamura, K. Shikano, and H. Kuwabara, "Voice conversion through vector quantization," in *Proc. ICASSP88*, pp. 655–658, 1988.



H. Valbret, E. Mulines, and J. Tubach, "Voice transformation using PSOLA techniques," *Speech Communication*, vol. 11, no. 2-3, pp. 175–187, 1992.



N. Iwahashi and Y. Sagisaka, "Speech spectrum transformation based on speaker interpolation," in *Proc. ICASSP94*, 1994.



Y. Stylianou, O. Cappé, and E. Moulines, "Statistical methods for voice quality transformation," *Proc. EUROSPEECH*, 1995.

Speech
Modifications

Yannis
Stylianou

Outline of the
talk

Modifications

Modeling and
Modifications

References



A. Kain and M. Macon, "Spectral voice conversion for text-to-speech synthesis," in *Proc. ICASSP98*, pp. 285–288, 1998.



T. Toda, H. Saruwatari, and K. Shikano, "Voice Conversion Algorithm based on Gaussian Mixture Model with Dynamic Frequency Warping of STRAIGHT spectrum," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Salt Lake City, USA), pp. 841–844, 2001.



T. Toda, A. Black, and K. Tokuda, "Spectral Conversion Based on Maximum Likelihood Estimation considering Global Variance of Converted Parameter," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Philadelphia, USA), pp. 9–12, 2005.



O. Turk and L. M. Arslan, "Robust processing techniques for voice conversion," *Computer Speech and Language*, vol. 20, pp. 441–467, 2006.



D. Sündermann, H. Hoega, A. Bonafonte, H. Ney, A. Black, and S. Narayanan, "Text-independent voice conversion based on unit selection," in *Proc. ICASSP06*, (Toulouse, France), pp. 81–84, 2006.



A. Mouchtaris, J. V. derSpiegel, and P. Mueller, "Non parallel training for voice conversion based on a parameter adaptation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 3, pp. 952–963, 2006.



D. Erro, *Intra-lingual and cross-lingual Voice Conversion using Harmonic plus Stochastic models*. PhD thesis, UPC, Barcelona, Spain, June 2008.

Speech
Modifications

**Yannis
Stylianou**

Outline of the
talk

Modifications

Modeling and
Modifications

References