Speech recognition and synthesis

Introduction

R&D in general purpose TTS systems is almost completely directed towards concatenative synthesis. Special purpose systems for speech research, visual speech generation, and small footprint applications still use Articulatory Synthesis or rule based systems (eg, [Boersma and Weenink(2004), Möhler(2005)])

Developping concatenative TTS systems [Black and Lenzo(2003a)]

- A strength is that it produces natural sounding speech from recorded human speech
- A weakness is that its quality totally depends on the qualities of the original recorded voice
- Evaluation must separate voice characteristics and system characteristics

Introduction

R&D in general purpose TTS systems is almost completely directed towards concatenative synthesis. Special purpose systems for speech research, visual speech generation, and small footprint applications still use Articulatory Synthesis or rule based systems (eg, [Boersma and Weenink(2004), Möhler(2005)])

Developping concatenative TTS systems [Black and Lenzo(2003a)]

- A strength is that it produces natural sounding speech from recorded human speech
- A weakness is that its quality totally depends on the qualities of the original recorded voice
- Evaluation must separate voice characteristics and system characteristics

Introduction

R&D in general purpose TTS systems is almost completely directed towards concatenative synthesis. Special purpose systems for speech research, visual speech generation, and small footprint applications still use Articulatory Synthesis or rule based systems (eg, [Boersma and Weenink(2004), Möhler(2005)])

Developping concatenative TTS systems [Black and Lenzo(2003a)]

- A strength is that it produces natural sounding speech from recorded human speech
- A weakness is that its quality totally depends on the qualities of the original recorded voice
- Evaluation must separate voice characteristics and system characteristics

Recording a voice: Speaker selection

Characteristics of a "good speaker"^{TM} [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"^{TM} [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"™ [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"^{TM} [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"™ [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"™ [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"™ [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"™ [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"™ [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Speaker selection

Characteristics of a "good speaker"™ [Black and Lenzo(2003a)]

- Availability and willingness (long recording times)
- Clear voice
- Consistent speaking (variability is bad)
- Will form the personality of the synthesis
- Will sign over all rights to you:
  - free for any use
  - free to distribute to anyone but cannot be used for commercial purposes without further contract.
  - research use only (does this allow public demos?)
  - fully proprietary
- Note: The style of speaking determines the style of the synthesis

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Diphone database

Diphone lists (≈1600 diphones)

- Choose phoneset
- Construct diphone list in nonsense words, eg [*pau t aa b aa b aa pau*]
- Add special or foreign phonemes and clusters
- Synthesize prompts as sounds for presentation
  - Text is ambiguous
  - Consistent prosody
  - Consistent pronunciation
- Record words under the best of circumstances
- Label and align phones (automatically)
- Extract pitch marks (electroglottogram)
- Build parameter files, eg, [MBROLA(2005)]
- Build and test database itself

Recording a voice: Unit database

Unit selection TTS is based on general speech, prosodic variation is good

- Size: phone, diphone, demi-syllable
- Type: phone, phone+stress, phone+word
- Concatenate units "in context", eg, stressed vs unstressed or word-initial vs -final phones
- Select units that fit requirements best
- Could use general speech corpus, but this generally lacks coverage and consistency
- Best to record a specially designed database

Recording a voice: Unit database

Unit selection TTS is based on general speech, prosodic variation is good

- Size: phone, diphone, demi-syllable
- Type: phone, phone+stress, phone+word
- Concatenate units "in context", eg, stressed vs unstressed or word-initial vs -final phones
- Select units that fit requirements best
- Could use general speech corpus, but this generally lacks coverage and consistency
- Best to record a specially designed database

Recording a voice: Unit database

Unit selection TTS is based on general speech, prosodic variation is good

- Size: phone, diphone, demi-syllable
- Type: phone, phone+stress, phone+word
- Concatenate units "in context", eg, stressed vs unstressed or word-initial vs -final phones
- Select units that fit requirements best
- Could use general speech corpus, but this generally lacks coverage and consistency
- Best to record a specially designed database

Recording a voice: Unit database

Unit selection TTS is based on general speech, prosodic variation is good

- Size: phone, diphone, demi-syllable
- Type: phone, phone+stress, phone+word
- Concatenate units "in context", eg, stressed vs unstressed or word-initial vs -final phones
- Select units that fit requirements best
- Could use general speech corpus, but this generally lacks coverage and consistency
- Best to record a specially designed database

Recording a voice: Unit database

Unit selection TTS is based on general speech, prosodic variation is good

- Size: phone, diphone, demi-syllable
- Type: phone, phone+stress, phone+word
- Concatenate units "in context", eg, stressed vs unstressed or word-initial vs -final phones
- Select units that fit requirements best
- Could use general speech corpus, but this generally lacks coverage and consistency
- Best to record a specially designed database

Recording a voice: Unit database

Unit selection TTS is based on general speech, prosodic variation is good

- Size: phone, diphone, demi-syllable
- Type: phone, phone+stress, phone+word
- Concatenate units "in context", eg, stressed vs unstressed or word-initial vs -final phones
- Select units that fit requirements best
- Could use general speech corpus, but this generally lacks coverage and consistency
- Best to record a specially designed database

Recording a voice: Constructing a unit database

Use a general language corpus with utterances that cover all relevant phenomena (Festival [Black and Lenzo(2003a)])

- Design the prompts (greedy algorithms)
- Record the prompts (best of circumstances)
- Autolabel the prompts
- Build utterance structures for recorded utterances
- Extract pitchmarks and build LPC coefficients (electroglottogram)
- Build a unit based synthesizer from the utterances
- Test and tune

Recording a voice: Constructing a unit database

Use a general language corpus with utterances that cover all relevant phenomena (Festival [Black and Lenzo(2003a)])

- Design the prompts (greedy algorithms)
- Record the prompts (best of circumstances)
- Autolabel the prompts
- Build utterance structures for recorded utterances
- Extract pitchmarks and build LPC coefficients (electroglottogram)
- Build a unit based synthesizer from the utterances
- Test and tune

Recording a voice: Constructing a unit database

Use a general language corpus with utterances that cover all relevant phenomena (Festival [Black and Lenzo(2003a)])

- Design the prompts (greedy algorithms)
- Record the prompts (best of circumstances)
- Autolabel the prompts
- Build utterance structures for recorded utterances
- Extract pitchmarks and build LPC coefficients (electroglottogram)
- Build a unit based synthesizer from the utterances
- Test and tune

Recording a voice: Constructing a unit database

Use a general language corpus with utterances that cover all relevant phenomena (Festival [Black and Lenzo(2003a)])

- Design the prompts (greedy algorithms)
- Record the prompts (best of circumstances)
- Autolabel the prompts
- Build utterance structures for recorded utterances
- Extract pitchmarks and build LPC coefficients (electroglottogram)
- Build a unit based synthesizer from the utterances
- Test and tune

Recording a voice: Constructing a unit database

Use a general language corpus with utterances that cover all relevant phenomena (Festival [Black and Lenzo(2003a)])

- Design the prompts (greedy algorithms)
- Record the prompts (best of circumstances)
- Autolabel the prompts
- Build utterance structures for recorded utterances
- Extract pitchmarks and build LPC coefficients (electroglottogram)
- Build a unit based synthesizer from the utterances
- Test and tune

Recording a voice: Constructing a unit database

Use a general language corpus with utterances that cover all relevant phenomena (Festival [Black and Lenzo(2003a)])

- Design the prompts (greedy algorithms)
- Record the prompts (best of circumstances)
- Autolabel the prompts
- Build utterance structures for recorded utterances
- Extract pitchmarks and build LPC coefficients (electroglottogram)
- Build a unit based synthesizer from the utterances
- Test and tune

Recording a voice: Constructing a unit database

Use a general language corpus with utterances that cover all relevant phenomena (Festival [Black and Lenzo(2003a)])

- Design the prompts (greedy algorithms)
- Record the prompts (best of circumstances)
- Autolabel the prompts
- Build utterance structures for recorded utterances
- Extract pitchmarks and build LPC coefficients (electroglottogram)
- Build a unit based synthesizer from the utterances
- Test and tune

Processing a voice: Autolabeling

Use the prompts to label and segment

- The prompts have known labeling and segmentation
- Align the prompts to the recordings, eg, dynamic time warping or forced ASR alignment
- Can even be done when synthesized prompts are from a TTS of a different language
- If segmentation goes wrong, verify by hand
- Determine syllable stress and sentence accent from prompt specification
- Feed labels into utterance structures etc.

Processing a voice: Autolabeling

Use the prompts to label and segment

- The prompts have known labeling and segmentation
- Align the prompts to the recordings, eg, dynamic time warping or forced ASR alignment
- Can even be done when synthesized prompts are from a TTS of a different language
- If segmentation goes wrong, verify by hand
- Determine syllable stress and sentence accent from prompt specification
- Feed labels into utterance structures etc.

Processing a voice: Autolabeling

Use the prompts to label and segment

- The prompts have known labeling and segmentation
- Align the prompts to the recordings, eg, dynamic time warping or forced ASR alignment
- Can even be done when synthesized prompts are from a TTS of a different language
- If segmentation goes wrong, verify by hand
- Determine syllable stress and sentence accent from prompt specification
- Feed labels into utterance structures etc.

Processing a voice: Autolabeling

Use the prompts to label and segment

- The prompts have known labeling and segmentation
- Align the prompts to the recordings, eg, dynamic time warping or forced ASR alignment
- Can even be done when synthesized prompts are from a TTS of a different language
- If segmentation goes wrong, verify by hand
- Determine syllable stress and sentence accent from prompt specification
- Feed labels into utterance structures etc.

Processing a voice: Autolabeling

Use the prompts to label and segment

- The prompts have known labeling and segmentation
- Align the prompts to the recordings, eg, dynamic time warping or forced ASR alignment
- Can even be done when synthesized prompts are from a TTS of a different language
- If segmentation goes wrong, verify by hand
- Determine syllable stress and sentence accent from prompt specification
- Feed labels into utterance structures etc.

Processing a voice: Autolabeling

Use the prompts to label and segment

- The prompts have known labeling and segmentation
- Align the prompts to the recordings, eg, dynamic time warping or forced ASR alignment
- Can even be done when synthesized prompts are from a TTS of a different language
- If segmentation goes wrong, verify by hand
- Determine syllable stress and sentence accent from prompt specification
- Feed labels into utterance structures etc.

Speech characteristics: Expressive speech

Consistent pronunciation means little expression. Add different styles (professional speaker/actor)

[Bulut et al.(2002)Bulut, Narayanan, and Syrdal]

- Use appropriate style for task, eg, news, weather, stories
- Message has more effect in correct emotional state
- Very important when working for children
- Basic states: anger, happiness, sadness and neutral
- Prosodic models must be specific for each emotional state

Speech characteristics: Expressive speech

Consistent pronunciation means little expression. Add different styles (professional speaker/actor)

[Bulut et al.(2002)Bulut, Narayanan, and Syrdal]

- Use appropriate style for task, eg, news, weather, stories
- Message has more effect in correct emotional state
- Very important when working for children
- Basic states: anger, happiness, sadness and neutral
- Prosodic models must be specific for each emotional state

Speech characteristics: Expressive speech

Consistent pronunciation means little expression. Add different styles (professional speaker/actor)

[Bulut et al.(2002)Bulut, Narayanan, and Syrdal]

- Use appropriate style for task, eg, news, weather, stories
- Message has more effect in correct emotional state
- Very important when working for children
- Basic states: anger, happiness, sadness and neutral
- Prosodic models must be specific for each emotional state

Speech characteristics: Expressive speech

Consistent pronunciation means little expression. Add different styles (professional speaker/actor)

[Bulut et al.(2002)Bulut, Narayanan, and Syrdal]

- Use appropriate style for task, eg, news, weather, stories
- Message has more effect in correct emotional state
- Very important when working for children
- Basic states: anger, happiness, sadness and neutral
- Prosodic models must be specific for each emotional state

Speech characteristics: Expressive speech

Consistent pronunciation means little expression. Add different styles (professional speaker/actor)

[Bulut et al.(2002)Bulut, Narayanan, and Syrdal]

- Use appropriate style for task, eg, news, weather, stories
- Message has more effect in correct emotional state
- Very important when working for children
- Basic states: anger, happiness, sadness and neutral
- Prosodic models must be specific for each emotional state

Speech characteristics: Changing speaker characteristics (not yet feasible)

Each different "voice" needs a separate speaker. Only what has been recorded can be spoken

- Change voice characteristics to create a different speaker, eg, man to woman to child (Praat allows this as a demo)
- Change voice to a different language variant or style
- Add new (level of) "expressiveness"
- Emotional state can be manipulated to some degree in prosody alone
- Techniques from rule based synthesis are needed to change complex traits, like stress and emotional states, reliably

Speech characteristics: Changing speaker characteristics (not yet feasible)

Each different "voice" needs a separate speaker. Only what has been recorded can be spoken

- Change voice characteristics to create a different speaker, eg, man to woman to child (Praat allows this as a demo)
- Change voice to a different language variant or style
- Add new (level of) "expressiveness"
- Emotional state can be manipulated to some degree in prosody alone
- Techniques from rule based synthesis are needed to change complex traits, like stress and emotional states, reliably

Speech characteristics: Changing speaker characteristics (not yet feasible)

Each different "voice" needs a separate speaker. Only what has been recorded can be spoken

- Change voice characteristics to create a different speaker, eg, man to woman to child (Praat allows this as a demo)
- Change voice to a different language variant or style
- Add new (level of) "expressiveness"
- Emotional state can be manipulated to some degree in prosody alone
- Techniques from rule based synthesis are needed to change complex traits, like stress and emotional states, reliably

Speech characteristics: Changing speaker characteristics (not yet feasible)

Each different "voice" needs a separate speaker. Only what has been recorded can be spoken

- Change voice characteristics to create a different speaker, eg, man to woman to child (Praat allows this as a demo)
- Change voice to a different language variant or style
- Add new (level of) "expressiveness"
- Emotional state can be manipulated to some degree in prosody alone
- Techniques from rule based synthesis are needed to change complex traits, like stress and emotional states, reliably

Speech characteristics: Changing speaker characteristics (not yet feasible)

Each different "voice" needs a separate speaker. Only what has been recorded can be spoken

- Change voice characteristics to create a different speaker, eg, man to woman to child (Praat allows this as a demo)
- Change voice to a different language variant or style
- Add new (level of) "expressiveness"
- Emotional state can be manipulated to some degree in prosody alone
- Techniques from rule based synthesis are needed to change complex traits, like stress and emotional states, reliably

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation

Evaluation is the mother of progress

- Evaluate modules separately
- Construct rigorous and uniform evaluation procedures and criteria
- Separate diagnostic tests from full system evaluations
- Compare different system
- Standardize external input: Voice, texts, use
- TTS is evaluated by listeners
  - Self selected volunteers (eg, internet)
  - Paid naive listeners (eg, students)
  - Paid target groups (eg, office workers, K12 children)
  - TTS developers (Tit-for-Tat evaluation)
  - External Experts

Evaluation: Criteria

What can be evaluated (full system evaluation)

- Intelligibility at phoneme and word level
- Naturalness and pleasantness
- Intonation and prosody
- Stress positions and breaks
- Long text rendering (eg, intonation variation)
- Task appropriateness, ie, correct style
- Voice and style selection in multi-speaker story telling (when feasible)

Evaluation: Criteria

What can be evaluated (full system evaluation)

- Intelligibility at phoneme and word level
- Naturalness and pleasantness
- Intonation and prosody
- Stress positions and breaks
- Long text rendering (eg, intonation variation)
- Task appropriateness, ie, correct style
- Voice and style selection in multi-speaker story telling (when feasible)

Evaluation: Criteria

What can be evaluated (full system evaluation)

- Intelligibility at phoneme and word level
- Naturalness and pleasantness
- Intonation and prosody
- Stress positions and breaks
- Long text rendering (eg, intonation variation)
- Task appropriateness, ie, correct style
- Voice and style selection in multi-speaker story telling (when feasible)

Evaluation: Criteria

What can be evaluated (full system evaluation)

- Intelligibility at phoneme and word level
- Naturalness and pleasantness
- Intonation and prosody
- Stress positions and breaks
- Long text rendering (eg, intonation variation)
- Task appropriateness, ie, correct style
- Voice and style selection in multi-speaker story telling (when feasible)

Evaluation: Criteria

What can be evaluated (full system evaluation)

- Intelligibility at phoneme and word level
- Naturalness and pleasantness
- Intonation and prosody
- Stress positions and breaks
- Long text rendering (eg, intonation variation)
- Task appropriateness, ie, correct style
- Voice and style selection in multi-speaker story telling (when feasible)

Evaluation: Criteria

What can be evaluated (full system evaluation)

- Intelligibility at phoneme and word level
- Naturalness and pleasantness
- Intonation and prosody
- Stress positions and breaks
- Long text rendering (eg, intonation variation)
- Task appropriateness, ie, correct style
- Voice and style selection in multi-speaker story telling (when feasible)

Evaluation: Criteria

What can be evaluated (full system evaluation)

- Intelligibility at phoneme and word level
- Naturalness and pleasantness
- Intonation and prosody
- Stress positions and breaks
- Long text rendering (eg, intonation variation)
- Task appropriateness, ie, correct style
- Voice and style selection in multi-speaker story telling (when feasible)

Blizzard challenge 2005

Aim: Find better synthesis techniques by comparing systems on the same data

Evaluating corpus-based speech synthesis on common datasets [Black and Tokuda(2005), Bennett(2005)]

- Effort to start international comparative evaluation of TTS systems
- Which approaches work, which don't
- Distribute common unit database, generate full TTS system within two weeks
- Evaluate common texts, 250 sentences from 5 genres
- Prevent "cheating" where needed

Blizzard challenge 2005

Aim: Find better synthesis techniques by comparing systems on the same data

Evaluating corpus-based speech synthesis on common datasets [Black and Tokuda(2005), Bennett(2005)]

- Effort to start international comparative evaluation of TTS systems
- Which approaches work, which don't
- Distribute common unit database, generate full TTS system within two weeks
- Evaluate common texts, 250 sentences from 5 genres
- Prevent "cheating" where needed

Blizzard challenge 2005

Aim: Find better synthesis techniques by comparing systems on the same data

Evaluating corpus-based speech synthesis on common datasets [Black and Tokuda(2005), Bennett(2005)]

- Effort to start international comparative evaluation of TTS systems
- Which approaches work, which don't
- Distribute common unit database, generate full TTS system within two weeks
- Evaluate common texts, 250 sentences from 5 genres
- Prevent "cheating" where needed

Blizzard challenge 2005

Aim: Find better synthesis techniques by comparing systems on the same data

Evaluating corpus-based speech synthesis on common datasets [Black and Tokuda(2005), Bennett(2005)]

- Effort to start international comparative evaluation of TTS systems
- Which approaches work, which don't
- Distribute common unit database, generate full TTS system within two weeks
- Evaluate common texts, 250 sentences from 5 genres
- Prevent "cheating" where needed

Blizzard challenge 2005

Aim: Find better synthesis techniques by comparing systems on the same data

Evaluating corpus-based speech synthesis on common datasets [Black and Tokuda(2005), Bennett(2005)]

- Effort to start international comparative evaluation of TTS systems
- Which approaches work, which don't
- Distribute common unit database, generate full TTS system within two weeks
- Evaluate common texts, 250 sentences from 5 genres
- Prevent "cheating" where needed

Blizzard challenge 2005: Speech

Common speech databases

- CMU ARCTIC databases: 2 old + 2 new voices
- 1200 phonetically balanced sentences (5-15 words)
- Project Gutenberg novels (prose style) [Project Gutenberg(2005)]
- All words in CMUDICT [Festvox(2005)]
- Eg, *They were three hundred yards apart.*
- Automatically segmented and labeled [Festvox(2005)]

Blizzard challenge 2005: Speech

Common speech databases

- CMU ARCTIC databases: 2 old + 2 new voices
- 1200 phonetically balanced sentences (5-15 words)
- Project Gutenberg novels (prose style) [Project Gutenberg(2005)]
- All words in CMUDICT [Festvox(2005)]
- Eg, *They were three hundred yards apart.*
- Automatically segmented and labeled [Festvox(2005)]

Blizzard challenge 2005: Speech

Common speech databases

- CMU ARCTIC databases: 2 old + 2 new voices
- 1200 phonetically balanced sentences (5-15 words)
- Project Gutenberg novels (prose style) [Project Gutenberg(2005)]
- All words in CMUDICT [Festvox(2005)]
- Eg, *They were three hundred yards apart.*
- Automatically segmented and labeled [Festvox(2005)]

Blizzard challenge 2005: Speech

Common speech databases

- CMU ARCTIC databases: 2 old + 2 new voices
- 1200 phonetically balanced sentences (5-15 words)
- Project Gutenberg novels (prose style) [Project Gutenberg(2005)]
- All words in CMUDICT [Festvox(2005)]
- Eg, *They were three hundred yards apart.*
- Automatically segmented and labeled [Festvox(2005)]

Blizzard challenge 2005: Speech

Common speech databases

- CMU ARCTIC databases: 2 old + 2 new voices
- 1200 phonetically balanced sentences (5-15 words)
- Project Gutenberg novels (prose style) [Project Gutenberg(2005)]
- All words in CMUDICT [Festvox(2005)]
- Eg, *They were three hundred yards apart.*
- Automatically segmented and labeled [Festvox(2005)]

Blizzard challenge 2005: Speech

Common speech databases

- CMU ARCTIC databases: 2 old + 2 new voices
- 1200 phonetically balanced sentences (5-15 words)
- Project Gutenberg novels (prose style) [Project Gutenberg(2005)]
- All words in CMUDICT [Festvox(2005)]
- Eg, *They were three hundred yards apart.*
- Automatically segmented and labeled [Festvox(2005)]

Blizzard challenge 2005: Evaluation

5 text genres, 50 sentences each

- Novels, same stories as original sentences
  Joe Garland lives like a good fellow.
- News, standard press-wire
  The two countries agreed to resolve any conflict through . . . Interfax said.
- Conversation, human side of spoken dialog system
  Yeah, I guess it will and something downtown please.
- Phonetically confusable sentences
  Now we will say cold/colt again.
- Semantically unpredictable sentences (SUS)
  The unsure steaks overcame the zippy rudder.

Blizzard challenge 2005: Evaluation

5 text genres, 50 sentences each

- Novels, same stories as original sentences
  Joe Garland lives like a good fellow.
- News, standard press-wire
  The two countries agreed to resolve any conflict through . . . Interfax said.
- Conversation, human side of spoken dialog system
  Yeah, I guess it will and something downtown please.
- Phonetically confusable sentences
  Now we will say cold/colt again.
- Semantically unpredictable sentences (SUS)
  The unsure steaks overcame the zippy rudder.

Blizzard challenge 2005: Evaluation

5 text genres, 50 sentences each

- Novels, same stories as original sentences
  Joe Garland lives like a good fellow.
- News, standard press-wire
  The two countries agreed to resolve any conflict through . . . Interfax said.
- Conversation, human side of spoken dialog system
  Yeah, I guess it will and something downtown please.
- Phonetically confusable sentences
  Now we will say cold/colt again.
- Semantically unpredictable sentences (SUS)
  The unsure steaks overcame the zippy rudder.

Blizzard challenge 2005: Evaluation

5 text genres, 50 sentences each

- Novels, same stories as original sentences
  Joe Garland lives like a good fellow.
- News, standard press-wire
  The two countries agreed to resolve any conflict through . . . Interfax said.
- Conversation, human side of spoken dialog system
  Yeah, I guess it will and something downtown please.
- Phonetically confusable sentences
  Now we will say cold/colt again.
- Semantically unpredictable sentences (SUS)
  The unsure steaks overcame the zippy rudder.

Blizzard challenge 2005: Evaluation

5 text genres, 50 sentences each

- Novels, same stories as original sentences
  Joe Garland lives like a good fellow.
- News, standard press-wire
  The two countries agreed to resolve any conflict through . . . Interfax said.
- Conversation, human side of spoken dialog system
  Yeah, I guess it will and something downtown please.
- Phonetically confusable sentences
  Now we will say cold/colt again.
- Semantically unpredictable sentences (SUS)
  The unsure steaks overcame the zippy rudder.

Blizzard challenge 2005: Listeners

Listener groups (and number who completed all tests)

[Bennett(2005)]

- Speech experts, each participant provided 10 local experts (50)
- Volunteers over the web (60, unpaid)
- US undergraduates (58, paid)
- It proved to be difficult to get enough listeners (≈ 100)

Blizzard challenge 2005: Listeners

Listener groups (and number who completed all tests)

[Bennett(2005)]

- Speech experts, each participant provided 10 local experts (50)
- Volunteers over the web (60, unpaid)
- US undergraduates (58, paid)
- It proved to be difficult to get enough listeners (≈ 100)

Blizzard challenge 2005: Listeners

Listener groups (and number who completed all tests)

[Bennett(2005)]

- Speech experts, each participant provided 10 local experts (50)
- Volunteers over the web (60, unpaid)
- US undergraduates (58, paid)
- It proved to be difficult to get enough listeners (≈ 100)

Blizzard challenge 2005: Listeners

Listener groups (and number who completed all tests)

[Bennett(2005)]

- Speech experts, each participant provided 10 local experts (50)
- Volunteers over the web (60, unpaid)
- US undergraduates (58, paid)
- It proved to be difficult to get enough listeners (≈ 100)

Blizzard challenge 2005: Test types

Test types

Mean opinion scores on a five point scale for *Novels*, *News*, and *Conversation* and Word Error Rate for *Phonetically Confusable* and *Semantically Unpredictable Sentences*
[Bennett(2005)]

Assignment: Week 6

Locate speech synthesis sites on the web

- Listen to TTS examples
- Locate sites where you can enter your own text
- Select real sentences from the web, try out "difficult" ones
- Compare them to Nextens and Festival
- What goes well? What doesn't?
- Describe the differences.

Assignment: Week 6

Locate speech synthesis sites on the web

- Listen to TTS examples
- Locate sites where you can enter your own text
- Select real sentences from the web, try out "difficult" ones
- Compare them to Nextens and Festival
- What goes well? What doesn't?
- Describe the differences.

Assignment: Week 6

Locate speech synthesis sites on the web

- Listen to TTS examples
- Locate sites where you can enter your own text
- Select real sentences from the web, try out "difficult" ones
- Compare them to Nextens and Festival
- What goes well? What doesn't?
- Describe the differences.

Assignment: Week 6

Locate speech synthesis sites on the web

- Listen to TTS examples
- Locate sites where you can enter your own text
- Select real sentences from the web, try out "difficult" ones
- Compare them to Nextens and Festival
- What goes well? What doesn't?
- Describe the differences.

Assignment: Week 6

Locate speech synthesis sites on the web

- Listen to TTS examples
- Locate sites where you can enter your own text
- Select real sentences from the web, try out "difficult" ones
- Compare them to Nextens and Festival
- What goes well? What doesn't?
- Describe the differences.

Assignment: Week 6

Locate speech synthesis sites on the web

- Listen to TTS examples
- Locate sites where you can enter your own text
- Select real sentences from the web, try out "difficult" ones
- Compare them to Nextens and Festival
- What goes well? What doesn't?
- Describe the differences.

Further Reading I

Christina L. Bennett.

Large Scale Evaluation of Corpus-based Synthesizers: Results and Lessons from the Blizzard Challenge 2005.
In *Proceedings of Interspeech 2005, Lisboa, Portugal*, September 2005.
URL `http://festvox.org/blizzard/bc2005/IS052023.PDF`.

Alan W. Black and Kevin A. Lenzo.

Building Synthetic Voices.
Festvox, 2 January 2003a.
URL `http://festvox.org/bsv/`.
Published on the festvox website.

Alan W. Black and Kevin A. Lenzo.

Festvox.
Web, 2003b.
URL `http://festvox.org/`.
Festival speech synthesis.

Alan W. Black and Keiichi Tokuda.

The Blizzard Challenge 2005: Evaluating corpus-based speech synthesis on common datasets.
In *Proceedings of Interspeech 2005, Lisboa, Portugal*, September 2005.
URL `http://festvox.org/blizzard/bc2005/IS051946.PDF`.

P. Boersma.

Praat, a system for doing phonetics by computer.
Glot International, 5:341–345, 2001.
URL `http://www.Praat.org/`.

Further Reading II

P. Boersma and D. Weenink.

Praat 4.2: doing phonetics by computer.
Computer program: http://www.Praat.org/, 2004.
URL http://www.Praat.org/.

Paulus Petrus Gerardus Boersma.

Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives.
PhD thesis, University of Amsterdam, September 1998.
URL http://www.fon.hum.uva.nl/paul/papers/funphon.pdf.

Murtaza Bulut, Shrikanth S. Narayanan, and Ann K. Syrdal.

Expressive speech synthesis using a concatenative synthesizer.
In *Proceedings of ICSLP 2002, Denver, COLORADO*, September 2002.
URL http://www.research.att.com/projects/tts/papers/2002_ICSLP/expressive.pdf.

Ronald A. Cole, Joseph Mariani, Hans Uszkoreit, Annie Zaenen, and Victor Zue, editors.

Survey of the State of the Art in Human Language Technology.
Cambridge University Press, 1996.
URL http://cslu.cse.ogi.edu/HLTsurvey/.
ISBN 0-521-59277-1.

Festvox.

Festvox.
Web, 2005.
URL http://www.festvox.org/.

FSF.

GNU General Public License.
Web, June 1991.
URL http://www.gnu.org/licenses/gpl.html.

Further Reading III

MBROLA.
The MBROLA Project.
Web, 2005.
URL http://tcts.fpms.ac.be/synthesis/.
Synthesis.

Bernd Möbius.
word and syllable models for german text-to-speech synthesis.
In Mike Edgington, editor, *Third ESCA/COCOSDA Workshop on SPEECH SYNTHESIS*, 26 November 1998.
URL http://www.slt.atr.co.jp/cocosda/jenolan/Proc/r06/r06.pdf.

Gregor Möhler.
Examples of Synthesized Speech.
Web, 2005.
URL http://www.ims.uni-stuttgart.de/~moehler/synthspeech/.
Good web-site with many examples.

Nextens.
NeXTeNS: Open Source Text-to-Speech for Dutch.
Web, 2003.
URL http://nextens.uvt.nl/index.html.

Project Gutenberg.
Project gutenberg free ebook library.
Web, 2005.
URL http://www.gutenberg.org/.

Further Reading IV

Richard Sproat.
ECE 598: Sp eech Synthesis.
Web.
URL http://catarina.ai.uiuc.edu/ECE598/Lectures/klattlpc.pdf.

SRL.
Synthesis of Speech.
Web.
URL http://wagstaff.asel.udel.edu/speech/tutorials/synthesis/.
Speech Research Lab, A.I. duPont hospital for children and University of Delaware.

Appendix A

Copyright License

The GNU General Public License I

Version 2, June 1991

51 Franklin Street, Fifth Floor, Boston, MA 02110-1301, USA

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software—to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

The GNU General Public License II

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.
Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.
The precise terms and conditions for copying, distribution and modification follow.

Terms and Conditions For Copying, Distribution and Modification

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".
Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

The GNU General Public License III

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.
   You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.
2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:
   1. You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.
   2. You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.
   3. If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

The GNU General Public License IV

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.
Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.
In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

    1. Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
    2. Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
    3. Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The GNU General Public License V

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.
If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.
5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.
6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

The GNU General Public License VI

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.
   If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.
   It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.
   This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.
8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

The GNU General Public License VII

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.
Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

No Warranty

11. Because the program is licensed free of charge, there is no warranty for the program, to the extent permitted by applicable law. Except when otherwise stated in writing the copyright holders and/or other parties provide the program "as is" without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The entire risk as to the quality and performance of the program is with you. Should the program prove defective, you assume the cost of all necessary servicing, repair or correction.

The GNU General Public License VIII

12. In no event unless required by applicable law or agreed to in writing will any copyright holder, or any other party who may modify and/or redistribute the program as permitted above, be liable to you for damages, including any general, special, incidental or consequential damages arising out of the use or inability to use the program (including but not limited to loss of data or data being rendered inaccurate or losses sustained by you or third parties or a failure of the program to operate with any other programs), even if such holder or other party has been advised of the possibility of such damages.

End of Terms and Conditions

The GNU General Public License IX

Appendix: How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

> *one line to give the program's name and a brief idea of what it does.*
> *Copyright (C) yyyy name of author*
> *This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.*
>
> *This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.*
>
> *You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301, USA.*

Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

> *Gnomovision version 69, Copyright (C) yyyy name of author*
> *Gnomovision comes with ABSOLUTELY NO WARRANTY; for details type 'show w'.*
> *This is free software, and you are welcome to redistribute it under certain conditions; type 'show c' for details.*

The GNU General Public License X

The hypothetical commands `show w` and `show c` should show the appropriate parts of the General Public License. Of course, the commands you use may be called something other than `show w` and `show c`; they could even be mouse-clicks or menu items—whatever suits your program.
You should also get your employer (if you work as a programmer) or your school, if any, to sign a "copyright disclaimer" for the program, if necessary. Here is a sample; alter the names:

> *Yoyodyne, Inc., hereby disclaims all copyright interest in the program*
> *'Gnomovision' (which makes passes at compilers) written by James Hacker.*
> *signature of Ty Coon, 1 April 1989*
> *Ty Coon, President of Vice*

This General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Library General Public License instead of this License.