Evolutionary Process Indicators for Active iGAs Applied to Weight Tuning in Unit Selection TTS Synthesis

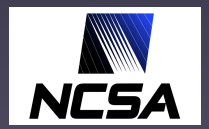
2010 IEEE Conference on Evolutionary Computation (Barcelona)

Lluís Formiga, Francesc Alías and Xavier Llorà

Media Technologies Research Group (GTM) - Universitat Ramon Llull.
C/Quatre Camins 2, 08022 Barcelona, Spain
{llformiga,falias}@salle.URL.edu

National Center for Supercomputing Applications (NCSA) - University of Illinois at Urbana-Champaign. 1205 W. Clark Street, Urbana IL 61801, USA
xllora@illinois.edu

July 21, 2010



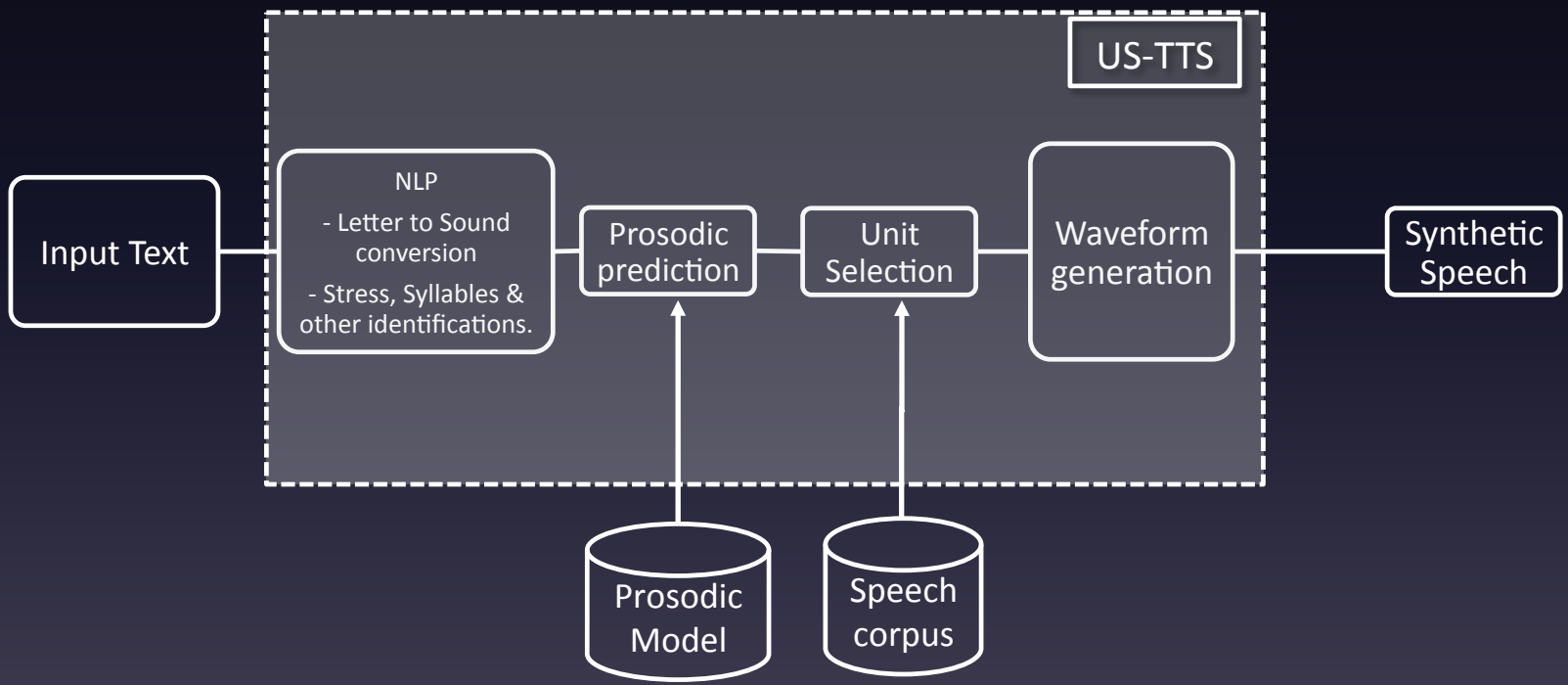
Index

Unit Selection Text-to-Speech (US-TTS)

- Text-to-Speech (TTS) systems produce speech from an input text.

Unit Selection Text-to-Speech (US-TTS)

- Text-to-Speech (TTS) systems produce speech from an input text.



- Unit Selection TTS (US-TTS) approach retrieves the best set of speech units from a previously recorded speech database.

Cost Function on US-TTS

- The retrieval method [Hunt and Black, 1996] is a Viterbi algorithm cost minimization based on a global cost function (equations 1,2,3).

$$C_T^i = \sum_{j=0}^{param\_t} w_j^i \cdot SC_T^i \quad (1)$$

$$C_C^i = \sum_{j=0}^{param\_c} w_j^i \cdot SC_C^i \quad (2)$$

$$C^i = C_T^i + C_C^i \quad (3)$$

where, w_j^i refers to the weight for the weighted subcost SC, which values are in the range of [0,1], considering $\sum w_i = 1$.

Cost Function on US-TTS

- The retrieval method [Hunt and Black, 1996] is a Viterbi algorithm cost minimization based on a global cost function (equations 1,2,3).

$$C_T^i = \sum_{j=0}^{param\_t} w_j^i \cdot SC_T^i \quad (1)$$

$$C_C^i = \sum_{j=0}^{param\_c} w_j^i \cdot SC_C^i \quad (2)$$

$$C^i = C_T^i + C_C^i \quad (3)$$

where, w_j^i refers to the weight for the weighted subcost SC, which values are in the range of [0,1], considering $\sum w_i = 1$.

- Ongoing discussion on scientific community:
  - Using only linguistic or acoustic subcosts on C_T^i
  - Which are the best concatenation subcosts (spectral, LSF…)
  - Units all share same weight pattern or it depends on unit type (context, stress, phonetic specs. …)

Open Issues on Weight Tuning

- On cutting edge non-evolutionary perceptual approaches:
  - Unique solution for each specifity is assumed (Unimodal).
  - Monosyllabic words don't represent real case.
  - Complexity relegated to user: $\approx$ 5 files evaluated simultaneously.
  - No fatigue, convergence, frustration or consistency indicators are considered to assess the search. [Alías et al., 2006].

Open Issues on Weight Tuning

- On cutting edge non-evolutionary perceptual approaches:
  - Unique solution for each specifity is assumed (Unimodal).
  - Monosyllabic words don't represent real case.
  - Complexity relegated to user: $\approx$ 5 files evaluated simultaneously.
  - No fatigue, convergence, frustration or consistency indicators are considered to assess the search. [Alías et al., 2006].

- User feedback must be incorporated on the optimization, avoiding simple validation, as done in [Toda et al., 2006]

Open Issues on Weight Tuning

- On cutting edge non-evolutionary perceptual approaches:
  - Unique solution for each specifity is assumed (Unimodal).
  - Monosyllabic words don't represent real case.
  - Complexity relegated to user: $\approx$ 5 files evaluated simultaneously.
  - No fatigue, convergence, frustration or consistency indicators are considered to assess the search. [Alías et al., 2006].

- User feedback must be incorporated on the optimization, avoiding simple validation, as done in [Toda et al., 2006]

- *Main Goal:*
  i) Find suitable weight values without specific knowledge of the parameters involved in US.

  ii) Tuning method must deal to optimize different patterns.

aiGA-based weight tuning

- active iGA successfully fused human and computer efforts on speech optimization [Alías et al., 2006, Alm and Llorà, 2006]

aiGA-based weight tuning

- active iGA successfully fused human and computer efforts on speech optimization [Alías et al., 2006, Alm and Llorà, 2006]
- Synthetic Fitness Basis:
  - Binary tourn. scheme ($s = 2$) introduce a partial order among the solutions.

aiGA-based weight tuning

- active iGA successfully fused human and computer efforts on speech optimization [Alías et al., 2006, Alm and Llorà, 2006]
- Synthetic Fitness Basis:
  - Binary tourn. scheme ($s = 2$) introduce a partial order among the solutions.
  - Partial order represented as a graph $\mathcal{G} =< \mathcal{V}, \mathcal{E} >$[Llorà. et al., 2005].

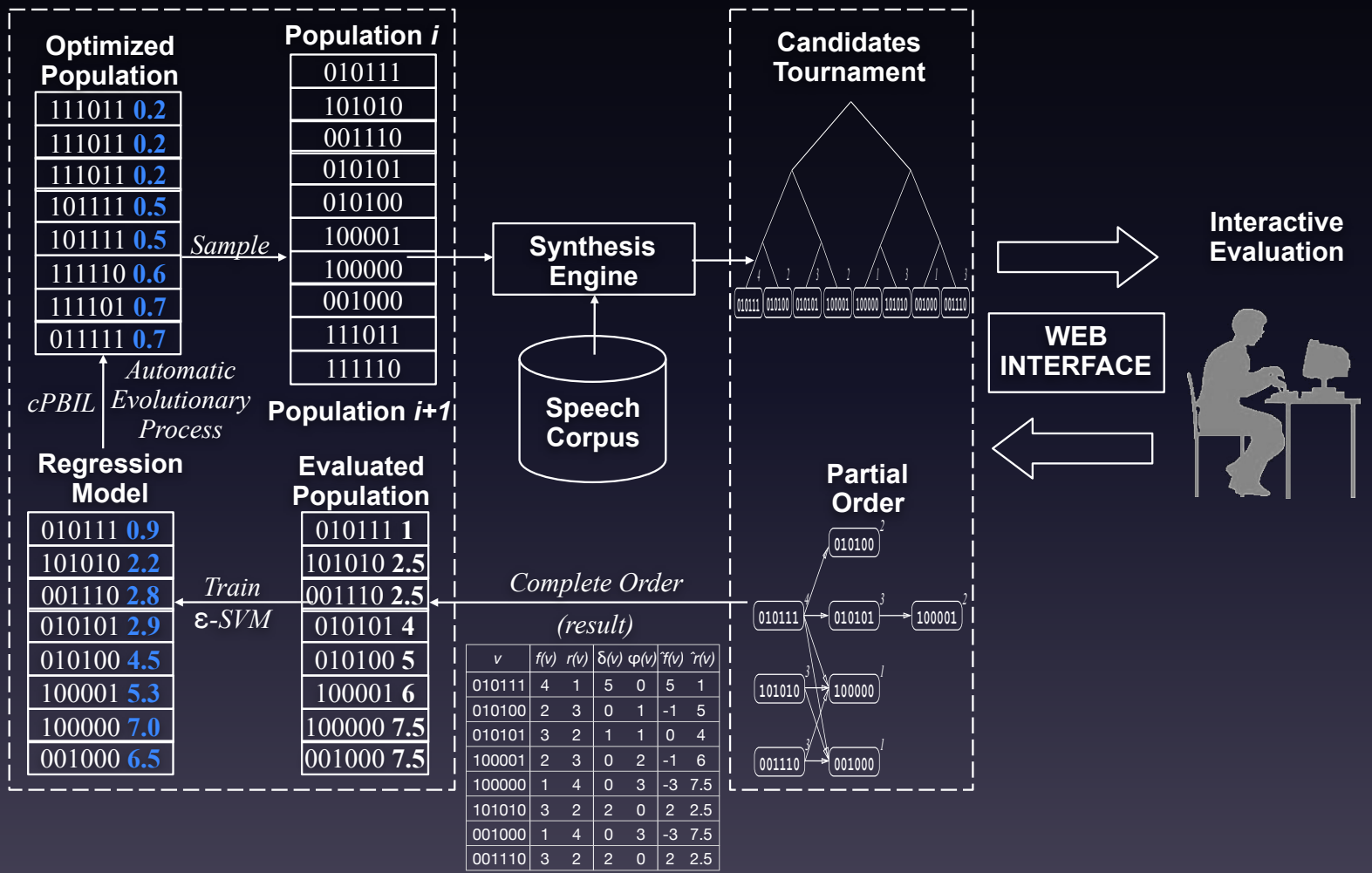
aiGA-based weight tuning

- active iGA successfully fused human and computer efforts on speech optimization [Alías et al., 2006, Alm and Llorà, 2006]
- Synthetic Fitness Basis:
  - Binary tourn. scheme ($s = 2$) introduce a partial order among the solutions.
  - Partial order represented as a graph $\mathcal{G} =< \mathcal{V}, \mathcal{E} >$[Llorà. et al., 2005].
  - Surrogate Fitness heuristics based on Pareto's dominance [Coello-Coello, 1998, Deb et al., 2000]

aiGA-based weight tuning

- active iGA successfully fused human and computer efforts on speech optimization [Alías et al., 2006, Alm and Llorà, 2006]
- Synthetic Fitness Basis:
  - Binary tourn. scheme ($s = 2$) introduce a partial order among the solutions.
  - Partial order represented as a graph $\mathcal{G} =< \mathcal{V}, \mathcal{E} >$[Llorà. et al., 2005].
  - Surrogate Fitness heuristics based on Pareto's dominance [Coello-Coello, 1998, Deb et al., 2000]
  - Global ordering measure computed using two dominance measures:
    i) $\delta(v)$: Number of vertexes recursively departing from v
    ii) $\phi(v)$: Number of vertexes recursively arriving to v

Method



v	f(v)	r(v)	δ(v)	φ(v)	f̂(v)	r̂(v)
010111	4	1	5	0	5	1
010100	2	3	0	1	-1	5
010101	3	2	1	1	0	4
100001	2	3	0	2	-1	6
100000	1	4	0	3	-3	7.5
101010	3	2	2	0	2	2.5
001000	1	4	0	3	-3	7.5
001110	3	2	2	0	2	2.5

Expanding the aiGAs: Evolutionary Process Indicators

- Reliability is a must on each kind of perceptual optimization.
- Indicators needed to assess the model obtained and the evolutionary process

Expanding the aiGAs: Evolutionary Process Indicators

- Reliability is a must on each kind of perceptual optimization.
- Indicators needed to assess the model obtained and the evolutionary process

- On [Alías et al., 2006] we introduced the κ-measure:
  - *i*) Considers cycles in the graph model as contradictions (e.g $A > B$, $B > C$, $C > A$) causing a decrease of the consistency of the model.
  - *ii*) κ-measure is computed as within-cycle/all vertexes ratio.
  - *iii*) Allows discarding bad evolutions from the result extraction process

Expanding the aiGAs: Evolutionary Process Indicators

- Reliability is a must on each kind of perceptual optimization.
- Indicators needed to assess the model obtained and the evolutionary process

- On [Alías et al., 2006] we introduced the κ-measure:
  - *i*) Considers cycles in the graph model as contradictions (e.g $A > B$, $B > C$, $C > A$) causing a decrease of the consistency of the model.
  - *ii*) κ-measure is computed as within-cycle/all vertexes ratio.
  - *iii*) Allows discarding bad evolutions from the result extraction process

- More indicators are needed for the following purposes:
  - *i*) Stop earlier the run due to noisy user evaluation (detect when he/she becomes fatigued/stucked): λ
  - *ii*) Discover convergence to unique/multiple solutions: ρ
  - *iii*) Ponder how user's agree with each other: τ

Certainty Ratio λ (1/2)

- λ-measure gives information about the confusion within the graph
- It considers equal evaluations (draws) as an increase of ambiguity.

Certainty Ratio λ (1/2)

- λ-measure gives information about the confusion within the graph
- It considers equal evaluations (draws) as an increase of ambiguity.
- Ambiguity may be caused by:
  - *i*) Premature Convergence of the Population
  - *ii*) Evaluator has become fatigued
- Defined as a within-cycle/all vertexes ratio.

Certainty Ratio λ (1/2)

- λ-measure gives information about the confusion within the graph
- It considers equal evaluations (draws) as an increase of ambiguity.
- Ambiguity may be caused by:
  - *i*) Premature Convergence of the Population
  - *ii*) Evaluator has become fatigued
- Defined as a within-cycle/all vertexes ratio.
- λ-measure also determines the effective test duration:
  - Once converged, all speech files seem indistinguishable to the evaluator
  - He/She tags them as equal, so no more building is required.
  - Direct consequences of low certainty ratio is a decrease on the ε-SVM regression accuracy.

Certainty Ratio λ (2/2)

- Models may stuck earlier and become noisy and ambiguous.

Certainty Ratio λ (2/2)

- Models may stuck earlier and become noisy and ambiguous.
- Stop-building point determined through λ-measure with a derivative averaging approach:

$$\begin{aligned} \Lambda &= \left[\lambda'_1, \ldots, \lambda'_t, \ldots, \lambda'_N\right], \quad \textit{where} \\ \lambda'_t &= \frac{\lambda(\mathcal{G}'^{t-1}) + \lambda(\mathcal{G}'^{t}) + \lambda(\mathcal{G}'^{t+1})}{3} \end{aligned}$$

$$sl_t = \lambda'_{t+1} - \lambda'_t, \forall t \in [1, \ldots, N-1]$$

$$t_{FINAL} = last(sl_t \geq 0)$$

where λ'_t is the averaged value $(\lambda_{t-1}, \lambda_t, \lambda_{t+1})$ of metric λ at iteration t. N is the total evaluations done by the user (45 in our case). Λ is the set of normalized λs and sl_t their derivative (slope).

Intra-user convergence ρ Inter-user correlation ratios τ

- *Intra-user* Convergence Ratio ρ
  - Measures convergence of the run either to a single solution or multiple solutions.
  - Final solution is obtained from best ranked individuals (aiGA has no replacement scheme).
  - Computed as the average of the correlation matrix (cosine correlation) among the best ranked solutions (our case 10% best).

Intra-user convergence ρ Inter-user correlation ratios τ

- *Intra-user* Convergence Ratio ρ
  - Measures convergence of the run either to a single solution or multiple solutions.
  - Final solution is obtained from best ranked individuals (aiGA has no replacement scheme).
  - Computed as the average of the correlation matrix (cosine correlation) among the best ranked solutions (our case 10% best).

Inter-user Correlation Ratio τ

- Gives information about of the similarity of each perceptual test performed by different users.
- Analogously to ρ-measure, we consider best ranked individuals.
- 10% best solution of each run are compared through a correlation matrix, which is averaged to obtain the final indicator.

Evolutionary Process Indicators

- Formally presented by equations:

$$\kappa\left(\mathcal{G}'^{t}, \omega_{\kappa}\right) = 1 - \left(\frac{1}{|\mathcal{V}'^{t}|} \cdot \sum_{v \in \chi(\mathcal{G}'^{t})} \omega_{v}^{\kappa}\right)^{\alpha_{\kappa}} \qquad \rho\left(\mathcal{G}'^{t}\right) = \left(\frac{1}{|\mathcal{B}'^{t}|} \cdot \sum_{\substack{v \in \mathcal{B} \\ vv \in \{\mathcal{B} | vv \neq v\}}} corr(v, vv)\right)^{\alpha_{\rho}}$$

$$\lambda\left(\mathcal{G}'^{t}, \omega_{\lambda}\right) = 1 - \left(\frac{1}{|\mathcal{V}'^{t}|} \cdot \sum_{v \in \psi(\mathcal{G}'^{t})} \omega_{v}^{\lambda}\right)^{\alpha_{\lambda}} \qquad \tau\left(\mathcal{G}'^{t}\right) = \left(\frac{1}{|\mathcal{U}'^{t}|} \cdot \sum_{\substack{v \in \mathcal{U} \\ vv \in \{\mathcal{U} | vv \neq v\}}} corr(v, vv)\right)^{\alpha_{\tau}}$$

Applicability of indicators:

Extraction of the Results context

- In [Llorà. et al., 2005] there were implemented two variants of extracting the final ranking table:

Applicability of indicators: Extraction of the Results context

- In [Llorà. et al., 2005] there were implemented two variants of extracting the final ranking table:
  - *i*) *Complete Order Dominance*: Considering all the vertexes on the graph pool
  - *ii*) *Longest Path Ranking*: Considering only the vertexes related to the longest dominance relation within the graph

Applicability of indicators: Extraction of the Results context

- In [Llorà. et al., 2005] there were implemented two variants of extracting the final ranking table:
  - *i*) *Complete Order Dominance*: Considering all the vertexes on the graph pool
  - *ii*) *Longest Path Ranking*: Considering only the vertexes related to the longest dominance relation within the graph

- Longest Path approach may be seen as an elitist approach: Avoids bad convergences. (Only one vertex is extracted as the best)

- Both methods will be applied to our experiments and analyzed.

Experiments

- [Alías et al., 2006] validated aiGA as extremely competitive for the US-TTS weight tuning task (proof-of-principle).

Experiments

- [Alías et al., 2006] validated aiGA as extremely competitive for the US-TTS weight tuning task (proof-of-principle).
- We want to test the suitability of aiGA when:
  - Complexity is increased
  - Bounding the maximum number of weights

Experiments

- [Alías et al., 2006] validated aiGA as extremely competitive for the US-TTS weight tuning task (proof-of-principle).
- We want to test the suitability of aiGA when:
  - Complexity is increased
  - Bounding the maximum number of weights
- New corpus: From 15 minutes of Catalan to 1 hour of Spanish recorded speech.

Experiments

- [Alías et al., 2006] validated aiGA as extremely competitive for the US-TTS weight tuning task (proof-of-principle).
- We want to test the suitability of aiGA when:
  - Complexity is increased
  - Bounding the maximum number of weights
- New corpus: From 15 minutes of Catalan to 1 hour of Spanish recorded speech.
- Subcosts expanded from 6 (acoustic) to 14 (acoustic / linguistic):
  i) Mismatches on Previous and Next Contexts of the unit, Part-Of-Speech, Position in utterance, Position in syllable, Position in word and Syllable Stress;
  ii) Target differences on half-phones durations, energy and pitch.
  iii) Energy, Mel-frequency cepstum and Pitch discontinuities at concatenation point.

Design of Experiments

- 16 Utterances (sentences) / 12 Evaluators

Design of Experiments

- 16 Utterances (sentences) / 12 Evaluators
  - 4 sentences assigned to each user
  - Assuring 3 users for each utterance
  - None of the users repeated the same 3 utterances set
  - Population size: 15 weight sets (individuals)([Llorà. et al., 2005]).
  - Maximum of 3 iterations [Alías et al., 2006]
  - Units of the utterance tagged as carrier (i.e. fixed) and eligible (i.e. variable) depending on its typology.

Design of Experiments

- 16 Utterances (sentences) / 12 Evaluators
  - 4 sentences assigned to each user
  - Assuring 3 users for each utterance
  - None of the users repeated the same 3 utterances set
  - Population size: 15 weight sets (individuals)([Llorà. et al., 2005]).
  - Maximum of 3 iterations [Alías et al., 2006]
  - Units of the utterance tagged as carrier (i.e. fixed) and eligible (i.e. variable) depending on its typology.
- Unit set selected according to:
  - Eligible sequence size (maximum run)
  - Linguistic Specificity of the Units (linguistic clustering)
  - Carrier/Eligible ratio (search width)
  - Number of different units to be selected (search depth)

Design of Experiments

- 16 Utterances (sentences) / 12 Evaluators
  - 4 sentences assigned to each user
  - Assuring 3 users for each utterance
  - None of the users repeated the same 3 utterances set
  - Population size: 15 weight sets (individuals)([Llorà. et al., 2005]).
  - Maximum of 3 iterations [Alías et al., 2006]
  - Units of the utterance tagged as carrier (i.e. fixed) and eligible (i.e. variable) depending on its typology.
- Unit set selected according to:
  - Eligible sequence size (maximum run)
  - Linguistic Specificity of the Units (linguistic clustering)
  - Carrier/Eligible ratio (search width)
  - Number of different units to be selected (search depth)
- After a aiGA run a graph model of 37 vertexes (weight combinations) is obtained

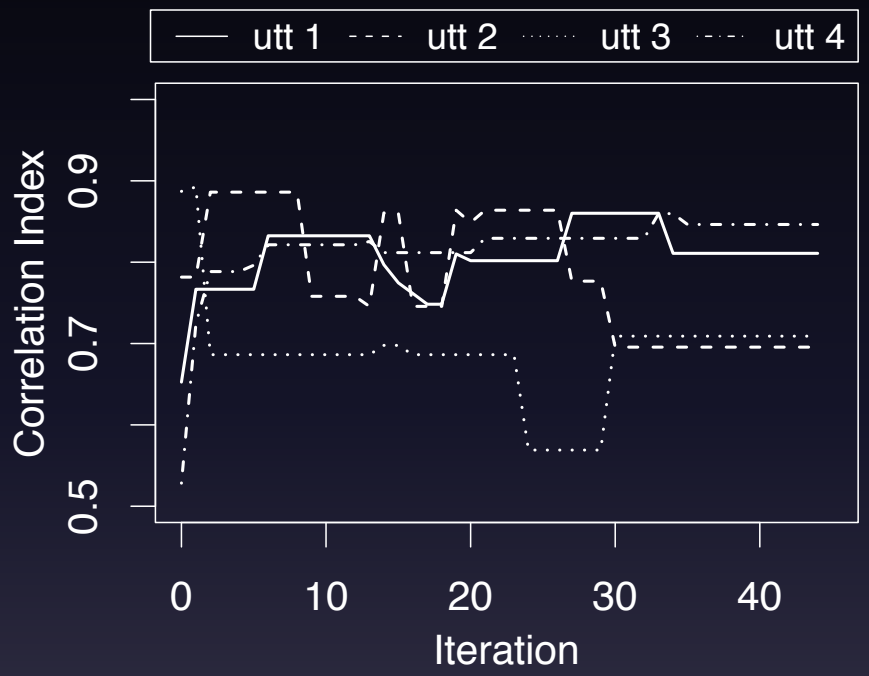
Indicators at work (1/2)

(a) *Evolution of consistency measure for an specific test utterance*

(b) *Evolution of certainty measure for an specific test utterance*

Indicators at work (2/2)

(c) *Evolution of intra-user convergence for an specific test utterance (longest path method)*



(d) *Evolution of inter-user correlation through four different test utterances (longest path method)*

Experiments considerations

- Median of each weight across different users winning solutions is considered as the final candidate weight value.

Experiments considerations

- Median of each weight across different users winning solutions is considered as the final candidate weight value.

- Obtaining Utterance patterns:
  - After obtaining candidates, we normalize ($\sum w_i = 1$)
  - Different user patterns are merged in terms of median computation.
  - We search actual user-evolved pattern most similar to the median pattern and consider it as the best.

Experiments considerations

- Median of each weight across different users winning solutions is considered as the final candidate weight value.

- Obtaining Utterance patterns:
  - After obtaining candidates, we normalize ($\sum w_i = 1$)
  - Different user patterns are merged in terms of median computation.
  - We search actual user-evolved pattern most similar to the median pattern and consider it as the best.

- Weight value differs depending on unit specification
- Utterance to unit mapping:
  i) All unit specifications in the evolved utterance adopt the weight values from the utterance which they were part of.
  ii) If they were part of more than one utterance, they adopt median values of all weight vectors related to mentioned utterances.

Experiments considerations

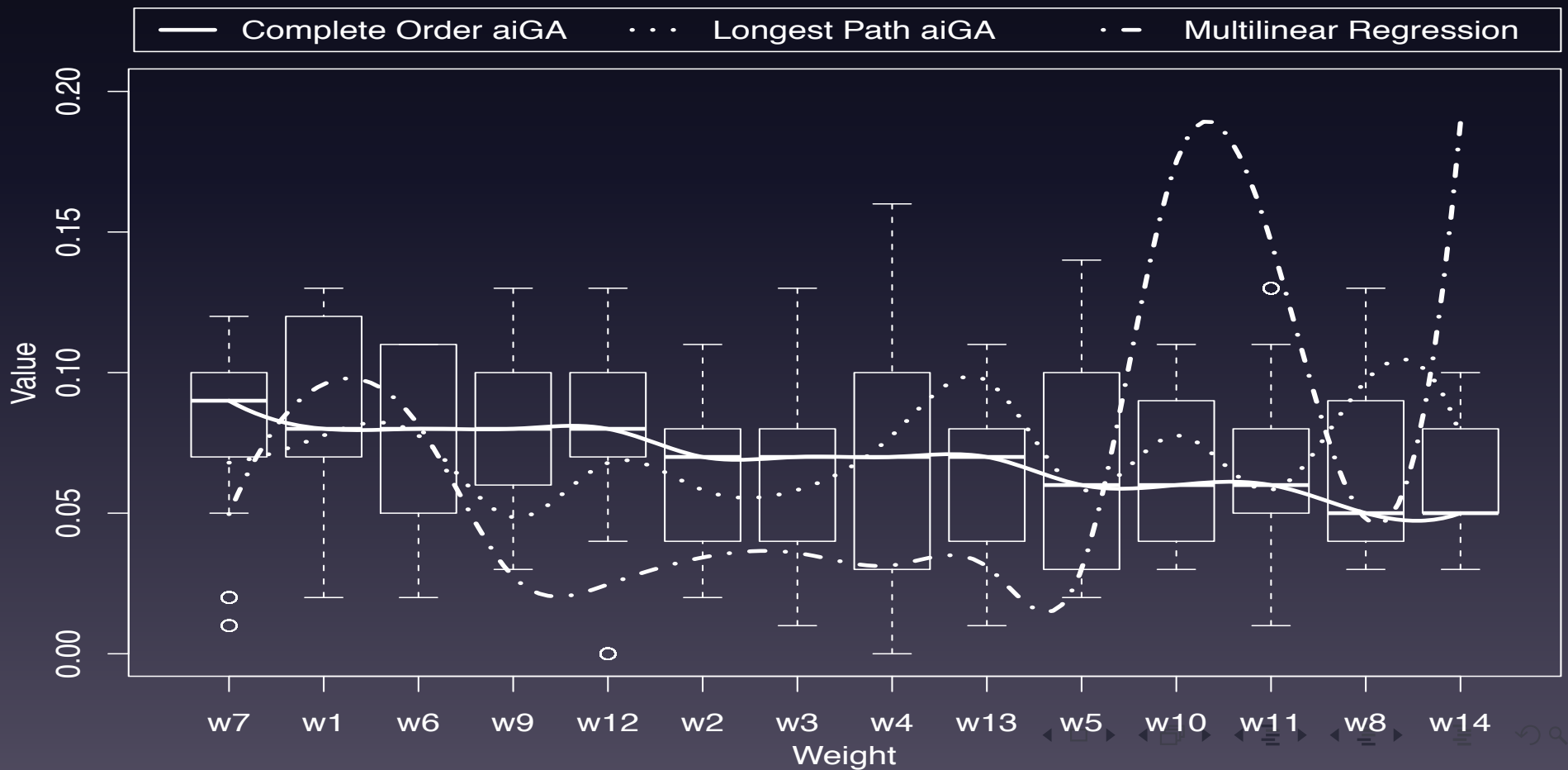
- Median of each weight across different users winning solutions is considered as the final candidate weight value.

- Obtaining Utterance patterns:
  - After obtaining candidates, we normalize ($\sum w_i = 1$)
  - Different user patterns are merged in terms of median computation.
  - We search actual user-evolved pattern most similar to the median pattern and consider it as the best.

- Weight value differs depending on unit specification
- Utterance to unit mapping:
  i) All unit specifications in the evolved utterance adopt the weight values from the utterance which they were part of.
  ii) If they were part of more than one utterance, they adopt median values of all weight vectors related to mentioned utterances.

Weights behaviour

- Methods compared: Methods compared: MLR/ aiGA (complete order/longest path)
- Patterns represent general behaviour through all units.



Experiments

- The proposed aiGA indicators are computed comparing longest path and complete order approaches (10 win / 2 tied / 4 lose), obtaining:

Test id	$\bar{\kappa}$	$\bar{\lambda}$	Complete Order		Longest Path	
			$\bar{\rho}$	$\bar{\tau}$	$\bar{\rho}$	$\bar{\tau}$
1	1	0.74	0.67	0.74	1	*0.81*
2	0.93	0.79	0.64	0.74	1	*0.77*
3	1	0.73	0.71	*0.71*	1	*0.71*
4	1	0.74	0.69	0.76	1	*0.85*
5	1	0.6	0.61	*0.78*	1	0.75
6	1	0.88	0.65	*0.78*	1	*0.78*
7	1	0.62	0.53	*0.72*	1	0.66
8	0.95	0.66	0.63	0.78	1	*0.80*
9	1	0.75	0.69	0.79	1	*0.86*
10	1	0.76	0.67	0.79	1	*0.87*
11	1	0.6	0.61	0.75	1	*0.82*
12	1	0.6	0.50	*0.71*	1	0.68
13	1	0.75	0.60	*0.71*	1	0.69
14	1	0.77	0.69	0.74	1	*0.76*
15	1	0.47	0.61	0.80	1	*0.95*
16	1	0.53	0.72	0.73	1	*0.82*

Results analysis

- More significant differences on Longest Path method than in Complete Order method, coherently with convergence measure.
- Subtle differences due to considering all prosodic specifications.

Results analysis

- More significant differences on Longest Path method than in Complete Order method, coherently with convergence measure.
- Subtle differences due to considering all prosodic specifications.
- PosInSyl, Duration, Conc. Energy, PosInUtterance, Prev. context are highly-weighted (accorging to complete-order-aiGA) although concretion depends on prosodic specification.

Results analysis

- More significant differences on Longest Path method than in Complete Order method, coherently with convergence measure.
- Subtle differences due to considering all prosodic specifications.
- PosInSyl, Duration, Conc. Energy, PosInUtterance, Prev. context are highly-weighted (accorging to complete-order-aiGA) although concretion depends on prosodic specification.
- aiGA is capable of combining acoustic and linguistic formulations on the cost function.

Results analysis

- More significant differences on Longest Path method than in Complete Order method, coherently with convergence measure.
- Subtle differences due to considering all prosodic specifications.
- PosInSyl, Duration, Conc. Energy, PosInUtterance, Prev. context are highly-weighted (accorging to complete-order-aiGA) although concretion depends on prosodic specification.
- aiGA is capable of combining acoustic and linguistic formulations on the cost function.
- Linguistics weights gain importance on subjective evolution. Otherwise on MLR (automatic regression) they achieve lower values.

Validation Stage (1/2)

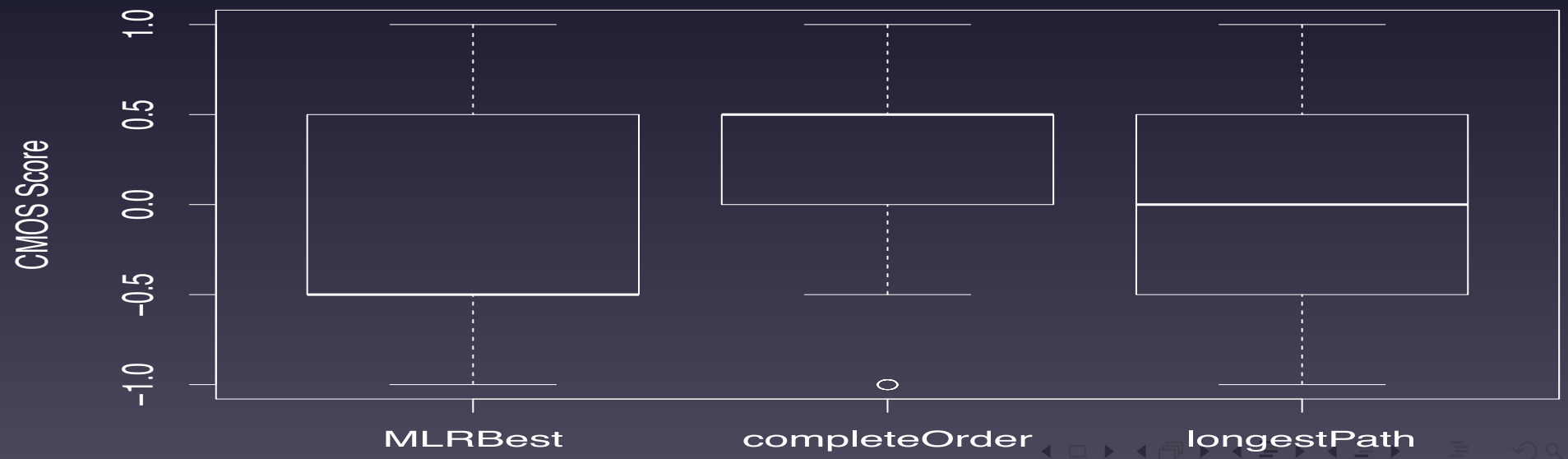
- 10 different sentences synthesized with:
  Methods compared: MLR/ aiGA (complete order/longest path)

Validation Stage (1/2)

- 10 different sentences synthesized with:
  Methods compared: MLR/ aiGA (complete order/longest path)
- Results compared using a 5 point CMOS across 13 users.
- On CMOS, two different synthesis of the same utterance are presented to the user and he/she is asked to choose between five election options: (definitely the first, the first, both, the second, definitely the second)

Validation Stage (1/2)

- 10 different sentences synthesized with:
  Methods compared: MLR/ aiGA (complete order/longest path)
- Results compared using a 5 point CMOS across 13 users.
- On CMOS, two different synthesis of the same utterance are presented to the user and he/she is asked to choose between five election options: (definitely the first, the first, both, the second, definitely the second)



Validation Stage (2/2)

- complete-order-aiGA > MLR / $p = 1.7 \cdot 10^{-5}$
- longest-path-aiGA > MLR / $p = 0.024$
- complete-order-aiGA > longest-path-aiGA / $p = 0.041$

Validation Stage (2/2)

- complete-order-aiGA > MLR / $p = 1.7 \cdot 10^{-5}$
- longest-path-aiGA > MLR / $p = 0.024$
- complete-order-aiGA > longest-path-aiGA / $p = 0.041$

- aiGA complete order is the best perceived method, despite the poor results on ρ and τ measures.
- MLR method (not perceptual) is perceived as the worst one.

Validation Stage (2/2)

- complete-order-aiGA > MLR / $p = 1.7 \cdot 10^{-5}$
- longest-path-aiGA > MLR / $p = 0.024$
- complete-order-aiGA > longest-path-aiGA / $p = 0.041$

- aiGA complete order is the best perceived method, despite the poor results on ρ and τ measures.
- MLR method (not perceptual) is perceived as the worst one.

- Better correlation/convergence indicators don't infer better results
- Diversity is preferred on building the final ranking of the graph despite being noisy and confusing (not elitist)

Conclusions (1/2)

- This work presents a next step on aiGA-based US-TTS weight tuning approach.

Conclusions (1/2)

- This work presents a next step on aiGA-based US-TTS weight tuning approach.
- Problem complexity increased:
  - Optimization variables extended from 7 acoustic to 14 acoustic/linguistic
  - New larger corpus being optimized

Conclusions (1/2)

- This work presents a next step on aiGA-based US-TTS weight tuning approach.
- Problem complexity increased:
  - Optimization variables extended from 7 acoustic to 14 acoustic/linguistic
  - New larger corpus being optimized
- New process indicators implemented beyond consistency measure κ in order to assess aiGA behavior across the run.
  - λ certainty ratio
  - ρ intra-user convergence ratio
  - τ inter-user correlation ratio

Conclusions (1/2)

- This work presents a next step on aiGA-based US-TTS weight tuning approach.
- Problem complexity increased:
  - Optimization variables extended from 7 acoustic to 14 acoustic/linguistic
  - New larger corpus being optimized
- New process indicators implemented beyond consistency measure κ in order to assess aiGA behavior across the run.
  - λ certainty ratio
  - ρ intra-user convergence ratio
  - τ inter-user correlation ratio
- Results validated by CMOS including baseline MLR.
- Validation stage states complete-order-aiGA as the best approach.

Conclusions (2/2)

- Counterintuitively, longest-path-aiGA approach does not yield to the best results even though adopting better correlation and convergence ratios.

Conclusions (2/2)

- Counterintuitively, longest-path-aiGA approach does not yield to the best results even though adopting better correlation and convergence ratios.
- Informal polls on [Llorà. et al., 2005] showed that the longest path method tended to be tedious and repetitive.

Conclusions (2/2)

- Counterintuitively, longest-path-aiGA approach does not yield to the best results even though adopting better correlation and convergence ratios.
- Informal polls on [Llorà. et al., 2005] showed that the longest path method tended to be tedious and repetitive.
- aiGA-based tuning method is capable of simultaneously tuning 14 weights

Conclusions (2/2)

- Counterintuitively, longest-path-aiGA approach does not yield to the best results even though adopting better correlation and convergence ratios.
- Informal polls on [Llorà. et al., 2005] showed that the longest path method tended to be tedious and repetitive.
- aiGA-based tuning method is capable of simultaneously tuning 14 weights
- In contrast to MLR, aiGA obtains reliable information over the linguistic/acoustic discussion without flattering linguistic weights (discrete) to zero values.

Conclusions (2/2)

- Counterintuitively, longest-path-aiGA approach does not yield to the best results even though adopting better correlation and convergence ratios.
- Informal polls on [Llorà. et al., 2005] showed that the longest path method tended to be tedious and repetitive.
- aiGA-based tuning method is capable of simultaneously tuning 14 weights
- In contrast to MLR, aiGA obtains reliable information over the linguistic/acoustic discussion without flattering linguistic weights (discrete) to zero values.
- Technical expertise of the evaluation users is not mandatory for conducting the tuning process

Acknowledgments

This work has been partially supported by the European Commission, Project SALERO (FP6 IST-4-027122-IP) and the Generalitat de Catalunya (grant 2009-SGR-293). We would like to thank the National Center for Supercomputing Applications for their support during the preparation of this manuscript.

Lluís Formiga and Francesc Alías
Media Technologies Group
Enginyeria la Salle
Ramon Llull University
{llformiga,falias}@salle.URL.edu

Xavier Llorà
NCSA
Univ. Illinois Urbana Champaign
xllora@illinois.edu

- Alías, F., Llorà, X., Formiga, L., Sastry, K., and Goldberg, D. E. (2006).
  Efficient interactive weight tuning for tts synthesis: reducing user fatigue by improving user consistency.
  In *Proceedings of ICASSP*, pages 865–868, Toulouse, France.
- Alías, F., Llorà, X., Iriondo, I., and Formiga, L. (2003).
  Ajuste subjetivo de pesos para selección de unidades a través de algoritmos genéticos interactivos.
  Procesamiento del Lenguaje Natural, 31:75–82.
- Alm, C. and Llorà, X. (2006).
  Evolving emotional prosody.
  In *Proc. of InterSpeech*, Pittsburgh.
- Coello-Coello, C. A. (December, 1998).
  An updated survey of ga-based multiobjective optimization techniques.
  Technical report, Laboratorio Nacional de Informática Avanzada (LANIA), Xalapa, Veracruz, Mexico.

type: Technical Report Lania-RD-09-08.

Coorman, G., Fackrell, J., Rutten, P., and Coile, B. V. (2000).
Segment selection in the l&h realspeak laboratory tts system.
In *Proc. of InterSpeech*, pages 395–398, Beijing, China.

Deb, K., Agrawal, S., Pratab, A., and Meyarivan, T. (2000).
A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii.
Technical report, Indian Institute of Technology.
type: KanGAL report 200001.

Hunt, A. and Black, A. W. (1996).
Unit selection in a concatenative speech synthesis system using a large speech database.
In *Proceedings of ICASSP*, pages 373–376, Atlanta, USA.

Iriondo, I., Socoro, J. C., and Alías, F. (2007).
Prosody modelling of spanish for expressive speech synthesis.
In *Proc. of InterSpeech*, volume 4, pages 821–824, Honolulu, Hawai'i, USA.

- Kaszczuk, M. and Osowski, L. (2009).
  The IVO Software Blizzard Challenge 2009 Entry: Improving IVONA Text-To-Speech.
  In *Proc. of InterSpeech*, Brighton UK.
- Lee, M., Lopresti, D., and Olive, J. (2003).
  A Text-to-Speech Platform for Variable Length Optimal Unit Searching using Perception Based Cost Functions.
  International Journal of Speech Technology, 6(4):347–356.
- Lee, M., Lopresti, D. P., and Olive, J. P. (2001).
  A text-to-speech platform for variable length optimal unit searching using perceptual cost functions.
  In *The 4th ISCA Workshop on Speech Synthesis*, pages 75–80, Perthshire, Scotland, UK.
- Llorà., X., Sastry, K., Goldberg, D. E., Gupta, A., and Lakshmi, L. (2005).
  Combating User Fatigue in iGAs: Partial Ordering, Support Vector Machines, and Synthetic Fitness.

Proceedings of Genetic and Evolutionary Computation Conference 2005 (GECCO-2005), pages 1363–1371.
note: (Also IlliGAL Report No. 2005009).


Meron, Y. and Hirose, K. (1999).
Efficient weight training for selection based synthesis.
In *Proceedings of EuroSpeech*, volume 5, pages 2319–2322, Budapest, Hungary.


Nelder, J. and Mead, R. (1965).
A simplex method for function minimization.
The computer journal, 7(4):308.


Park, S. S., Kim, C. K., and Kim, N. S. (2003).
Discriminative weight training for unit-selection based speech synthesis.
In *Proc. of InterSpeech*, volume 1, pages 281–284, Ginebra, Switzerland.


Schröder, M., Pammi, S., and Türk, O. (2009).
Multilingual MARY TTS participation in the Blizzard Challenge 2009.

In *Proc. of InterSpeech*, Brighton UK.

Sebag, M. and Ducoulombier, Q. (1998).
Extending Population-Based Incremental Learning to Continuous Search Spaces.
Lecture Notes in Computer Science, 1498:418–427.

Srinivas, N. and Deb, K. (1994).
Multiobjective optimization using nondominated sorting in genetic algorithms.
In *Journal on Evolutionary Computation, vol. 2, num. 3*.

Takagi, H. (2001).
Interactive evolutionary computation: fusion of the capabilities of the ec optimization and human evaluation.
Proceedings of the IEEE, 89(9):1275–1296.

Taylor, P. (2006).
The Target Cost Formulation in Unit Selection Speech Synthesis.
In *Proc. of InterSpeech*, Pittsburgh.

Toda, T., Kawai, H., Tsuzaki, M., and Shikano, K. (2006).

An evaluation of cost functions sensitively capturing local degradation of naturalness for segment selection in concatenative speech synthesis.
In *Speech Communication, Elsevier*, volume 48, pages 45–56.


Tsuzaki, M. (2001).
Feature extraction by auditory modeling for unit selection in concatenative speech synthesis.
In *Seventh European Conference on Speech Communication and Technology*.


Vapnik, V. N. (1999).
The Nature of Statistical Learning Theory.
Springer-Verlag New York, Inc.


Wouters, J. and Macon, M. (1998).
A perceptual evaluation of distance measures for concatenative speech synthesis.
In *in Proc. ICSLP*.