ECE Illinois

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN | COLLEGE OF ENGINEERING

Statistical Testing Of Off-line Comparative Subjective Evaluations For Optimizing Perceptual Conversational Quality In VoIP

Batu Sat and Benjamin W. Wah

12/15/2008

Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana-Champaign, USA

ECE Illinois

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Outline

- Introduction
- Approach & Problem Statement
- Model and Deductions of Subjective Comparisons
- Subjective Evaluations Methods
- Strategy for Simultaneous Evaluations
- Experimental Results

General Problem Studied

- Design the operation of control schemes
  - Real-time multi-media communication systems
  - Achieves high perceptual conversational quality
  - Robust to dynamic network conditions & communication scenarios
- Systems with common properties
  - Trade-offs among objective metrics on subjective preferences
  - Constrained resources on best-effort IP network
  - Communication scenario among participants

Subjective Evaluations

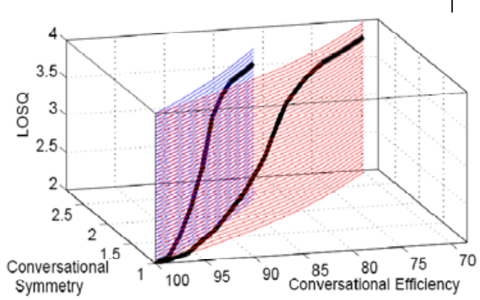
- On-line subjective evaluations are infeasible
  - Offline subjective tests are expensive and require multiple subjects
- Absolute Category Rating (e.g. ITU P.800 MOS)
  - Two operating points with multiple quality metrics may not be comparable
  - Not very accurate for small difference or high quality
  - Statistical significance cannot be associated with MOS differences
  - Suitable for verification of system performance

Outline

- Introduction
- Approach & Problem Statement
- Model and Deductions of Subjective Comparisons
- Subjective Evaluations Methods
- Strategy for Simultaneous Evaluations
- Experimental Results

Play-out Scheduling Design

- System control
  - MED: Trade-off between LOSQ and delay degradations
  - Goal: Choose optimal MED at run time – optimal operating point
- Network conditions: non-stationary & connection dependent
- Conversational scenarios
  - Frequency of conversation turns
  - Speech and silence durations
- Multiple quality metrics
  - Operating curve in multi-D space



Our Approach

- **Comparative ranking leading to partial order**
- **Dividing the problem into two stages**
  - Identify best operating point off-line given operating curve
  - Learn and generalize from limited number of conditions at run-time
- **Simulation & evaluation of results under given conditions**
  - Repeatability of subjective tests → relating results to control parameter
- **Pruning of search space**
  - Small changes in objective space may not be subjectively perceptible
  - Systematically use previous subjective preference results to reduce future tests
- **Combining of multiple pair-wise comparisons using Bayesian framework**
- **Learning of a classifier to generalize to similar but unseen conditions**

Problem Statement

- Statistical scheduling of off-line comparative subjective tests for evaluating alternative operating points on an operating curve of a control scheme in real-time multi-media systems
- Assumptions:
  - Domain knowledge on problem: identify monotonic quality metrics
  - Region of Dominance (ROD) is known on operating curve
- Not studied in this paper:
  - Multiple operating curves corresponding to different conditions
  - Learning and generalization of a classifier

ECE Illinois — DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Outline

ECE Illinois — DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Model of Subjective Comparisons

Notation:

- Operating curve: ***O*** , set of feasible points
- A^{min}, A^{max} : two extreme points ***O***
- Comparative Opinion Distribution when comparing A and B
  - COD(A,B) = (p_{-1}, p_0, p_1, p_2)

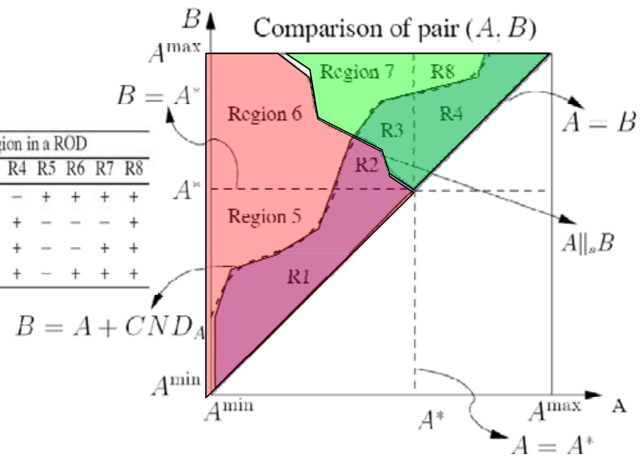
Condition	Probability	Notation
A is better than B	$Pr(A > B)$	$p_1(A, B)$
A is about the same as B	$Pr(A \approx B)$	$p_0(A, B)$
A is worse than B	$Pr(A < B)$	$p_{-1}(A, B)$
A is incomparable to B	$Pr(A?B)$	$p_2(A, B)$

Model of Subjective Comparisons

2-D representation of comparing A and B

- 4 boundary lines and 8 regions: based on relative location to A*

Boundary Line	Region in a ROD							
	R1	R2	R3	R4	R5	R6	R7	R8
$B-A-CND_A$	−	−	−	−	+	+	+	+
$A-A^*$	−	−	−	+	−	−	−	+
$A\|_s B$	−	−	+	+	−	−	+	+
$B-A^*$	−	+	+	+	−	+	+	+



Axioms of Subjective Comparisons (1/3)

Reflectivity

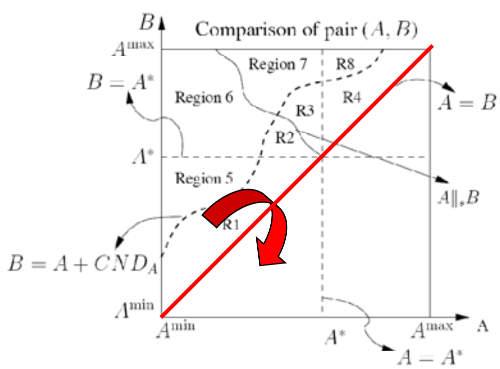
- Comparing a point with itself: $p_0(A,A) = 1$

Independent and Identically Distributed

- Finite no. of IID samples

Symmetry/anti-symmetry

- Order of comparison does not affect outcome
- $p_0(A,B) = p_0(B,A)$
- $p_2(A,B) = p_2(B,A)$
- $p_1(A,B) = p_{-1}(B,A)$

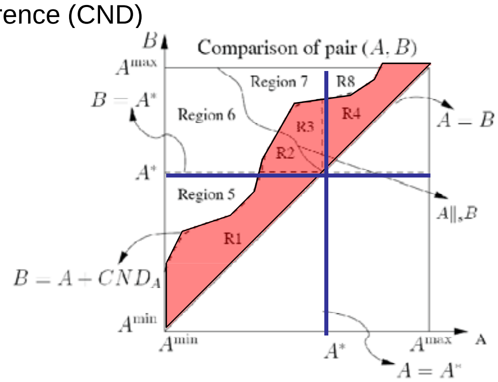


ECE Illinois — DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Axioms of Subjective Comparisons (2/3)

Just Noticeable Difference (JND)
- 50% of subjects perceiving a difference in quality with respect to A
- Complete Noticeable Difference (CND)

Indistinguishability
- $|B-A| <= JND_A$
- $p_0(A,B)$ is monotonically non-increasing w.r.t. B-A



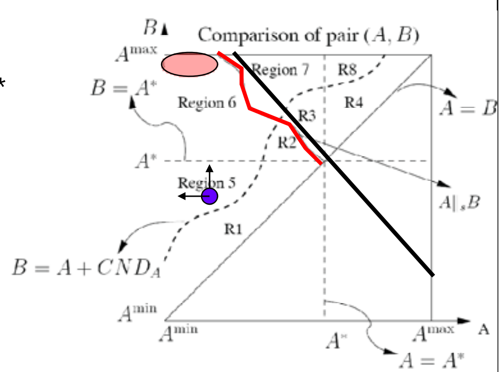
Locally optimal point
- $A^* = \{ A \mid p_1(A,B) > 0.5$ for all B in O such that $|B-A| > JND_A\}$
- A* is preferred among all points within the ROD, except within JND

ECE Illinois — DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Axioms of Subjective Comparisons (3/3)

Incomparability
- Perceptible degradations are different between A and B

Subjective Preference
- $|p_1(A,B) - p_{-1}(A,B)|$ increases as point closer to A* is perturbed towards A*



Control Symmetry
- $|A-A^*| = |B-A^*|$

Subjective Symmetry
- $p_1(A,B) = p_{-1}(A,B)$

ECE Illinois

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Deductions on Optimal Alternative

- Simplified parametric model needed
  - Allow information learned on multiple comparisons to be combined
- Belief function
  - Representing knowledge on location of A*
  - $f_{A^*}(\boldsymbol{a})$, where ***a*** on ***O***
- Initial Knowledge
  - Assuming uniformly distributed $f^0_{A^*}(a) = 1,\ a \in [A^{\min}, A^{\max}]$

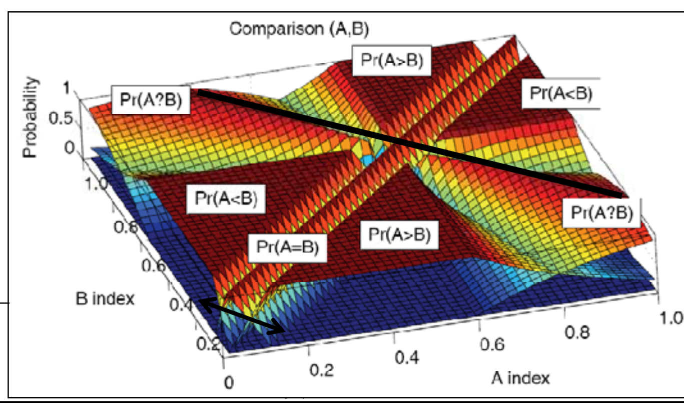
ECE Illinois

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Simplified Parametric Model (1/2)

Assumption 1: *CND* and *JND* are constant and do not vary with respect to *A* within *ROD* of local optimum

Assumption 2: Boundary line representing subjective symmetric pairs $A||_sB$ is a straight line



Simplified Parametric Model (2/2)

Assumption 3:

m and ***n*** are stochastic: $B = mA + n = \frac{-\gamma}{\Delta - \gamma} A + \frac{\Delta}{\Delta - \gamma} A^*$

$A||_sB$ is defined by red point (A*,A*) and one of green points

Green point is equally likely on $B - A = \Delta$ line thus,

γ is uniform in $[0, \Delta]$

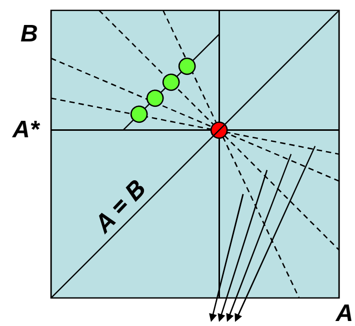
Assumption 4:

If A<A*<B and B > {B | $A||_sB$ },

then $p_{-1}(A,B) = 0$

If B < {B | $A||_sB$ } or {B | $A||_sB$ }

does not exist, then $p_1(A,B) = 0$



possible $A||_sB$ lines

Deductions on Single Pairwise Comparison

Bayesian Formulation

- Posterior probability from prior probability and new evidence

$$f_{A^*}(a|\overline{p}) = \frac{L(a|COD(A,B) = \overline{p}) * f_{A^*}(a)}{\int_0^1 L(\eta|COD(A,B) = \overline{p}) * f_{A^*}(\eta) d\eta}$$

- Likelihood function L(***a***|***p***) indicates the likelihood of obtaining ***p*** as the result of a subjective comparison of (A,B) if A* = ***a***
- Likelihood is obtained using occurrence frequencies of 4 outcomes
  - A >s B → (A,B) in regions 1,2,5 or 6 → any $a \in [A^{\min}, A + \gamma]$ can be A^*
  - A <s B → (A,B) in regions 3,4,7 or 8 → any $a \in [A + \gamma, A^{\max}]$ can be A^*
  - A =s B → (A,B) in regions 1,2,3 or 4 → No deduction
  - A ?s B → (A,B) in regions 1 through 8 → No deduction

ECE Illinois
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Deductions on Single Pairwise Comparison

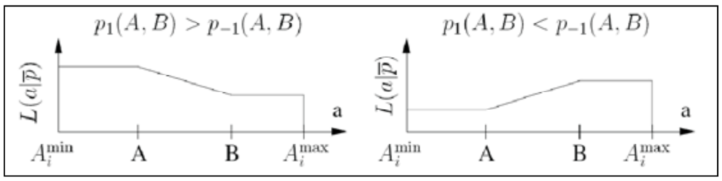
- Conditioned on the value of γ. Likelihood function is

$$L(a|\overline{p},\gamma) = \begin{cases} p_1 + p_0 + p_2 & \text{if } A^{\min} < a < A + \gamma \\ p_{-1} + p_0 + p_2 & \text{if } A + \gamma < a < A^{\max}. \end{cases}$$

- Unconditioned likelihood function is

$$L(a|\overline{p}) = E_\gamma[L(a|\overline{p},\gamma)] \;=\; \int_0^\Delta L(a|\overline{p},\gamma)Pr(\gamma)d\gamma$$

$$= \begin{cases} p_0 + p_2 + p_1 & \text{if } A^{\min} < a < A \\ p_0 + p_2 + \frac{p_1(B-a)+p_{-1}(a-A)}{B-A} & \text{if } A \le a \le B \\ p_0 + p_2 + p_{-1} & \text{if } B < a < A^{\max}. \end{cases}$$



ECE Illinois
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Deductions on Subsequent Comparisons

- Combined belief function after n[th] comparison;

$$f^n_{A^*}(a) = \frac{f^{n-1}_{A^*}(a) * L(a|COD(A_n,B_n)=\overline{p})}{\int_{A^{\min}}^{A^{\max}} f^{n-1}_{A^*}(\eta) * L(\eta|COD(A_n,B_n)=\overline{p})d\eta}.$$

- Combination is associative and in closed form

$$f^n_{A^*}(a) = \frac{\prod_{i=1}^n L(a|COD(A_n,B_n)=\overline{p})}{\int_{A^{\min}}^{A^{\max}} \prod_{i=1}^n L(\eta|COD(A_n,B_n)=\overline{p})d\eta}$$

Estimation of Optimal Alternative & Utility

- ***Utility*** of a belief function is the probability that A* estimate is within JND of A*
  - Estimation error of less than JND is insignificant
- ***A* estimate:*** operating point with maximum utility on ***O***

$$\hat{A}^*(f) = \arg\max_a \left\{ \int_{a-JND}^{a+JND} f(\xi)d\xi \right\}$$

- ***Utility***

$$U(f) = Pr(|\hat{A}^* - A^*| \leq JND) = \int_{\hat{A}^*-JND}^{\hat{A}^*+JND} f(\xi)d\xi$$

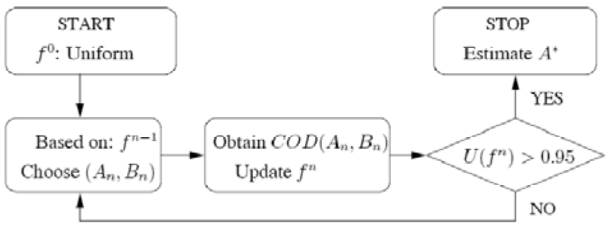
Outline

- Introduction
- Approach & Problem Statement
- Model of Subjective Comparisons
- Subjective Evaluations Methods
- Strategy for Simultaneous Evaluations
- Experimental Results

Subjective Evaluation Methods

Problem Formulation

- Find a set of M comparisons $\overline{A_n} = [A_n^1, \ldots, A_n^M]$ and $\overline{B_n} = [B_n^1, \ldots, B_n^M]$ in each batch, based on current belief function
- Choose $(\overline{A_n}, \overline{B_n})$ to min $n^* \doteq \min\{n \mid U(f^n) \geq 0.95\}$



- Simultaneous, Batch-based & Independent Evaluations

Strategy for Simultaneous Evaluations

- Minimize total comparisons by maximizing utility in each step:
  - S(U): expected number of comparisons left if U is the current utility

$$\begin{aligned} S(U(f^{n-1})) &= 1 + S(U(f^n)) \\ &= 1 + \min_{A_n, B_n} S(U(f^n | A_n, B_n)) \end{aligned}$$

- Observations on Bayesian Formulation
  - Likelihood function is unimodal with mode of A*
  - S(***U***) is a non-increasing function of ***U***
  - U(f^n) is a non-decreasing function of ***n***
- Comparison between (A,B) is most conclusive when $|p_1 - p_{-1}|$ is large
  - $\min\{p_0 + p_2\} \rightarrow \max\{|p_1 - p_{-1}|\}$
  - If A or B = A* and B-A = CND → $\{p_0 + p_2\}$ is minimized

Simultaneous and Batch Evaluations

- Optimal choice of next comparison pair (simultaneous)

$$(A_n, B_n) = \begin{cases} (\hat{A}^* - C\hat{N}D, \hat{A}^*) & \text{if } n \in Even \\ (\hat{A}^*, \hat{A}^* + C\hat{N}D) & \text{if } n \in Odd. \end{cases}$$

- Batch Based Evaluations (heuristic) – M per batch:

$$C^i = Mod(\frac{i-1}{M-1} + \hat{A}^*, 1)$$

$$(A_n^i, B_n^i) = \begin{cases} (C^i - C\hat{N}D, C^i) & \text{if } C^i < \hat{A}^* \\ (C^i, C^i + C\hat{N}D) & \text{if } C^i > \hat{A}^* \\ \text{Both pairs above} & \text{if } C^i = \hat{A}^* \end{cases}$$

Performance Analysis

Alg. 1: Complete pair-wise comparison among JND spaced points on ***O***
Alg. 2: Choose pairs randomly and uniformly in search space (any M)
Alg. 3: Choose a single pair (M=1) in each batch optimally
Alg. 4: Choose multiple (M>1) pairs in each batch using heuristic

Algorithm	$\bar{JND}$			
	0.1	0.03	0.01	0.003
1. Independent Eval.	45	≈ 500	≈ 5000	≈ 50000
2. Random (any M)	31.1	192	> 300	> 300
3. Optimal (M=1)	6.4	9.9	18.3	49.6
4. Heuristic (M=2)	6.7	11.3	21.4	56.5
Heuristic (M=3)	9.6	15.6	30.4	78.7
Heuristic (M=4)	14.0	19.6	34.2	81.2

Conclusion and Future Work

Conclusion

- Divide offline problem into two stages:
  - 1st given an operating curve find optimal operating point
  - 2nd learn/generalize multiple operating curves for run-time
- Bayesian framework to represent and combine information learned
- Adaptation of next comparison to reduce total comparisons
- Applicable to other real-time multimedia communication problems

Future Work

- Multiple local optima on one operating curve
- Multiple operating curves
- Learning/generalizing of classifiers