Sheep, Goats, Lambs and Wolves An Analysis of Individual Differences in Speaker Recognition Performance

11/23/98

These are the slides of the NIST coordinated presentation for the ICSLP '98 Conference in Sidney, Australia, November 1998. They are not part of the official Proceedings of this Conference.

Click here to start

Table of Contents

Author: George R. Doddington

Home Page: http://www.nist.gov/speech

Sheep, Goats, Lambs and Wolves

An Analysis of Individual Differences in Speaker Recognition Performance

George Doddington[1,2,3,5], Walter Liggett[1], Alvin Martin[1], Mark Przybocki[1], Douglas Reynolds[3,4]

[1] National Institute of Standards and Technology
[2] The Johns Hopkins University
[3] U.S. Department of Defense
[4] MIT Lincoln Laboratory
[5] SRI International

The Hypothetical Menagerie

- **Typical speakers:** The well-behaved majority.
  - ***Sheep***: Speakers who exhibit ***good true speaker acceptance***.
- **Problem speakers:** The troublesome minorities.
  - ***Goats***: Speakers who are *exceptionally* ***unsuccessful at being accepted***.
  - ***Lambs***: Speakers who are *exceptionally* ***vulnerable to impersonation by others***.
  - ***Wolves***: Speakers who are *exceptionally* ***successful at impersonating others***.

Conclusion: It's a zoo out there!

Goats, *Lambs* and *Wolves* exist!

Recommendation:

- Direct research toward the minimization of these undesirable animals.
- Include measurement of speaker differences as part of performance evaluation.

An html copy of this presentation is available at URL http://www.nist.gov/speech/icslp_98

Sheep, Goats, Lambs and Wolves An Analysis of Individual Differences in Speaker Recognition Performance

George Doddington1,2,3,5, Walter Liggett1, Alvin Martin1, Mark Przybocki1, Douglas Reynolds3,4

1 National Institute of Standards and Technology 2 The Johns Hopkins University 3 U.S. Department of Defense 4 MIT Lincoln Laboratory 5 SRI International

Next slide Back to first slide View graphic version

The Question

- Do these animals really exist? That is, do the speakers in the population of speakers truly exhibit ***individual differences***?

– *OR* –

- Are the observed performance differences merely a result of ***statistical variance*** (due to handset, message content, acoustic environment, etc.)?

The Hypothetical Menagerie

- **Typical speakers: The well-behaved majority.**
  - Sheep: Speakers who exhibit good true speaker acceptance.
- **Problem speakers: The troublesome minorities.**
  - Goats: Speakers who are exceptionally unsuccessful at being accepted.
  - Lambs: Speakers who are exceptionally vulnerable to impersonation by others.
  - Wolves: Speakers who are exceptionally successful at impersonating others.

Previous slide Next slide Back to first slide View graphic version

The Task and the Data

Text-independent speaker verification:
accept speaker if score ≥ threshold (if $x \geq \theta$).

- 30-second test segments (3400 segments, total)
- Conversational telephone speech
- 200 men and 220 women (*no cross-sex trials*)
- Different handsets used in training and test
- Data restricted to electret handsets

For details, refer to the ***1998 Speaker Recognition Evaluation Plan*** at URL ftp://jaguar.ncsl.nist.gov/evaluations/speaker/feb98/plans/current_plan.htm

The Question

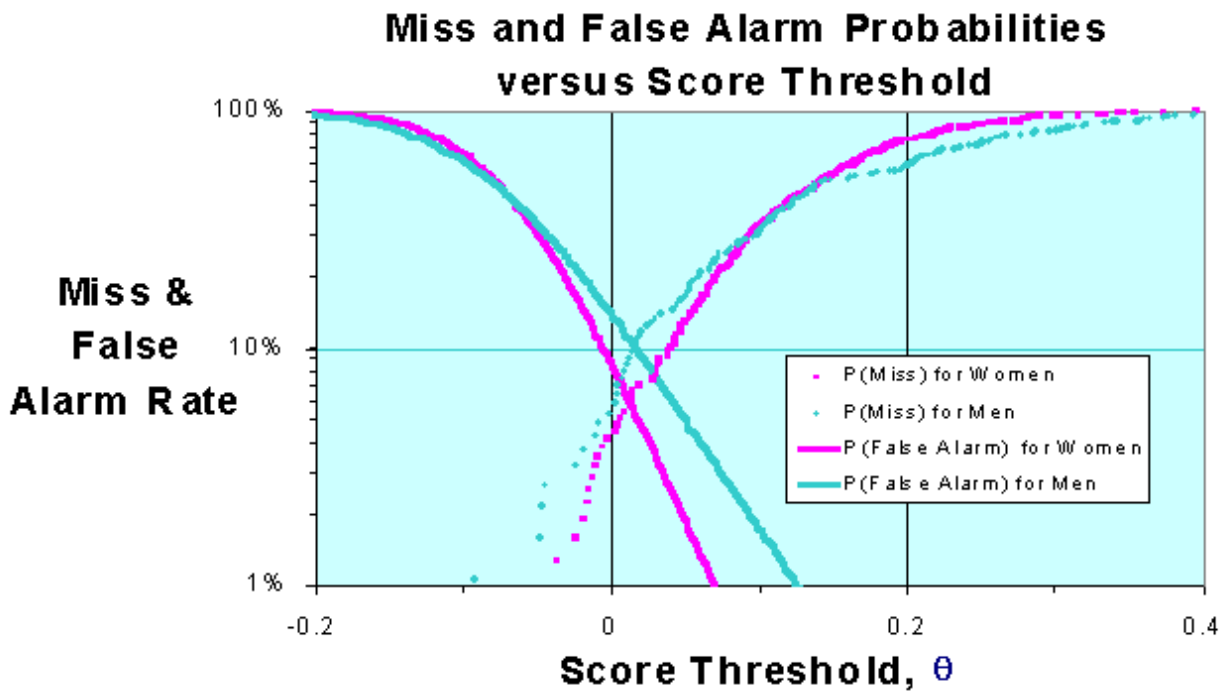
- **Do these animals really exist? That is, do the speakers in the population of speakers truly exhibit individual differences?**

  – OR –

- **Are the observed performance differences merely a result of statistical variance (due to handset, message content, acoustic environment, etc.)?**

Previous slide Next slide Back to first slide View graphic version

Speaker Recognition Performance



Slide 5 of 14

The Task and the Data

Text-independent speaker verification: accept speaker if score ³ threshold (if x ³ q).

- 30-second test segments (3400 segments, total)
- Conversational telephone speech
- 200 men and 220 women (no cross-sex trials)
- Different handsets used in training and test
- Data restricted to electret handsets

For details, refer to the 1998 Speaker Recognition Evaluation Plan at URL ftp://jaguar.ncsl.nist.gov/evaluations/speaker/feb98/plans/current_plan.htm

Previous slide Next slide Back to first slide View graphic version

The Goat Test Statistic

nonparametric Kruskal-Wallis rank sum test

- Each model speaker, M_i, contributes n_i test segments ($n_i \geq 5$).
- Each of these segments is scored against its respective model.

Model Speaker	Segment Index			
M_1	x_{11}	x_{12}	. . .	
M_2	x_{21}	x_{22}		
.	.			
.	.			
.	.			
M_i	x_{i1}	x_{i2}	rank sum	x_{in_i}
.	.			
.	.			
.	.			
M_m	x_{m1}	x_{m2}	. . .	

Test Procedure

- Replace each score by its rank among all scores: $x_{ij} \Rightarrow r_{ij}$.
- Sum the segment ranks for each model speaker: $R_i = \sum_{j=1}^{n_i} r_{ij}$
- Compute the Kruskal-Wallis test statistic:

$$H = \left\{ \frac{12}{N(N+1)} \cdot \sum_{i=1}^{m} \frac{R_i^2}{n_i} \right\} - 3(N+1)$$

where $N = \sum_{i=1}^{m} n_i$

Speaker Recognition Performance

Previous slide Next slide Back to first slide View graphic version

The Hypothetical Menagerie

- **Typical speakers:** The well-behaved majority.
  - ***Sheep***: A speaker who exhibits ***good true speaker acceptance***.
- **Problem speakers:** The troublesome minorities.
  - ***Goat***: A speaker who is *exceptionally* ***unsuccessful at being accepted***.
  - ***Lamb***: A ***model*** speaker who is *exceptionally* vulnerable to impersonation by others.
  - ***Wolf***: A ***segment*** speaker who is *exceptionally* ***successful at impersonating others***.

The Goat Test Statistic nonparametric Kruskal-Wallis rank sum test

- **Each model speaker, Mi, contributes ni test segments (ni ³ 5).**
- **Each of these segments is scored against its respective model.**

Test Procedure

- **Replace each score by its rank among all scores: xij Þ rij.**
- **Sum the segment ranks for each model speaker:**
- **Compute the Kruskal-Wallis test statistic:**

Previous slide Next slide Back to first slide View graphic version

Lamb/*Wolf* Test Statistics

- Each segment speaker, S_i, contributes n_i test segments
- Each of these segments is scored against all models

		Segment Speaker			
		S_1	S_2	. . .	S_s
Model Speaker	M_1	x_{11}	x_{12}	. . .	x_{1s}
	M_2	x_{21}	x_{22}	*Lamb*	x_{2s}
	⋮	⋮	*Wolf*		⋮
	M_m	x_{m1}	x_{m2}	. . .	x_{ms}

Test Procedure

- Compute average impostor scores, x_{ij}, for all S_i and M_j.
- Replace each score by its rank among all scores.
- Sum the ranks for each putative animal:
  - *Lamb*: **Row** sum
  - *Wolf*: **Column** sum

The Hypothetical Menagerie

- **Typical speakers: The well-behaved majority.**
  - Sheep: A speaker who exhibits good true speaker acceptance.
- **Problem speakers: The troublesome minorities.**
  - Goat: A speaker who is exceptionally unsuccessful at being accepted.
  - Lamb: A model speaker who is exceptionally vulnerable to impersonation by others.
  - Wolf: A segment speaker who is exceptionally successful at impersonating others.

Previous slide Next slide Back to first slide View graphic version

Statistical Significance

Goatish, lambish and wolfish behaviors were all demonstrated with ***confidence > 99%***

	The Kruskal-Wallis rank sum test statistic (χ^2)		
	D.F.	P = 0.01	Observed
Goat Test	68	73.3	168
Lamb Test	435	449.1	10253
Wolf Test	68	73.3	1782

Lamb/Wolf Test Statistics

- **Each segment speaker, Si, contributes ni test segments**
- **Each of these segments is scored against all models**

Test Procedure

- **Compute average impostor scores, xij, for all Si and Mj.**
- **Replace each score by its rank among all scores.**
- **Sum the ranks for each putative animal:**
  - Lamb: Row sum
  - Wolf: Column sum

Previous slide Next slide Back to first slide View graphic version

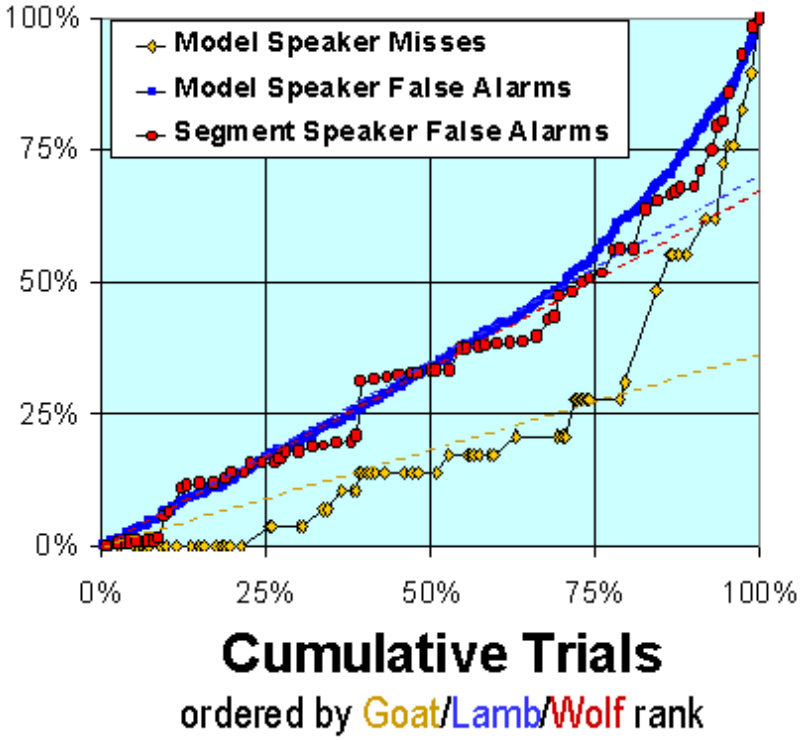
Distribution of Errors versus Animal Rankings

Cumulative Errors

Misses for Model Speakers

False Alarms for Model Speakers

False Alarms for Segment Speakers

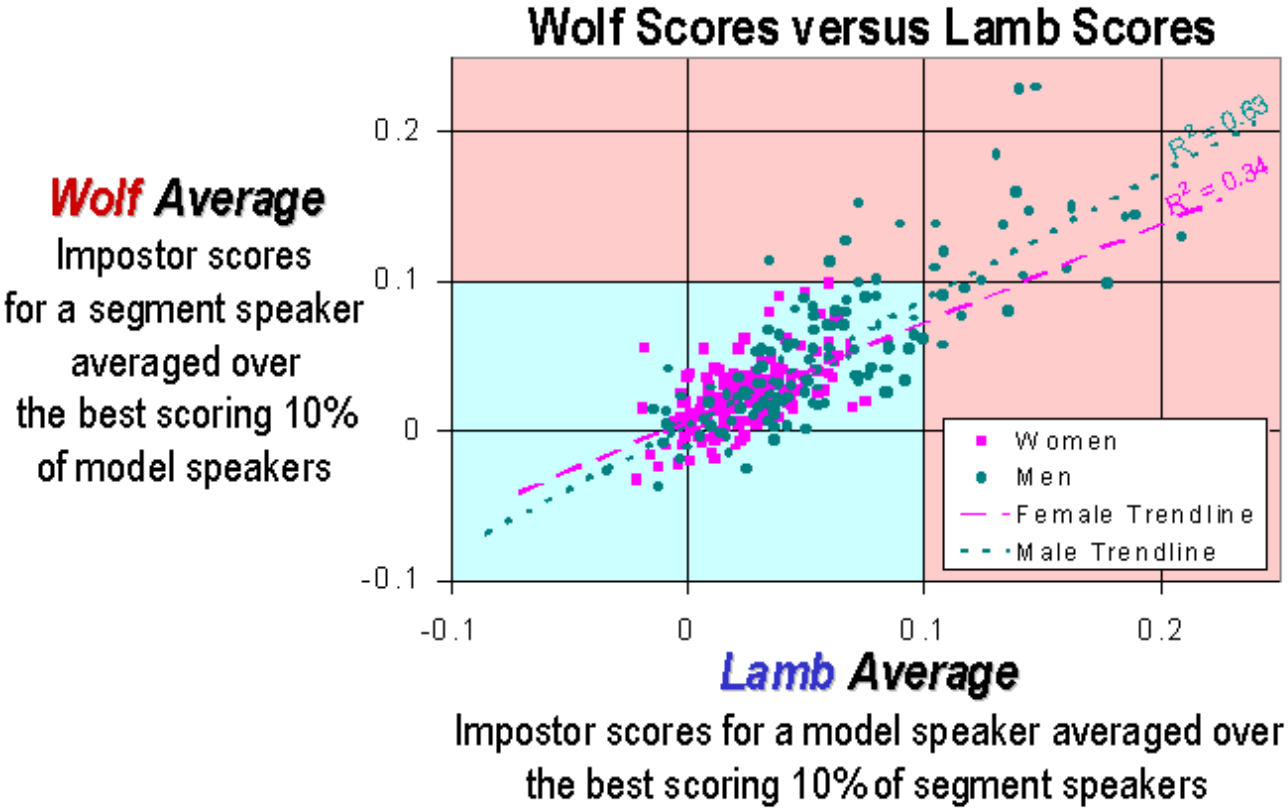


Statistical Significance

Goatish, lambish and wolfish behaviors were all demonstrated with confidence > 99%

Previous slide Next slide Back to first slide View graphic version

Can a Lamb be also a Wolf?



Distribution of Errors versus Animal Rankings

Cumulative Errors

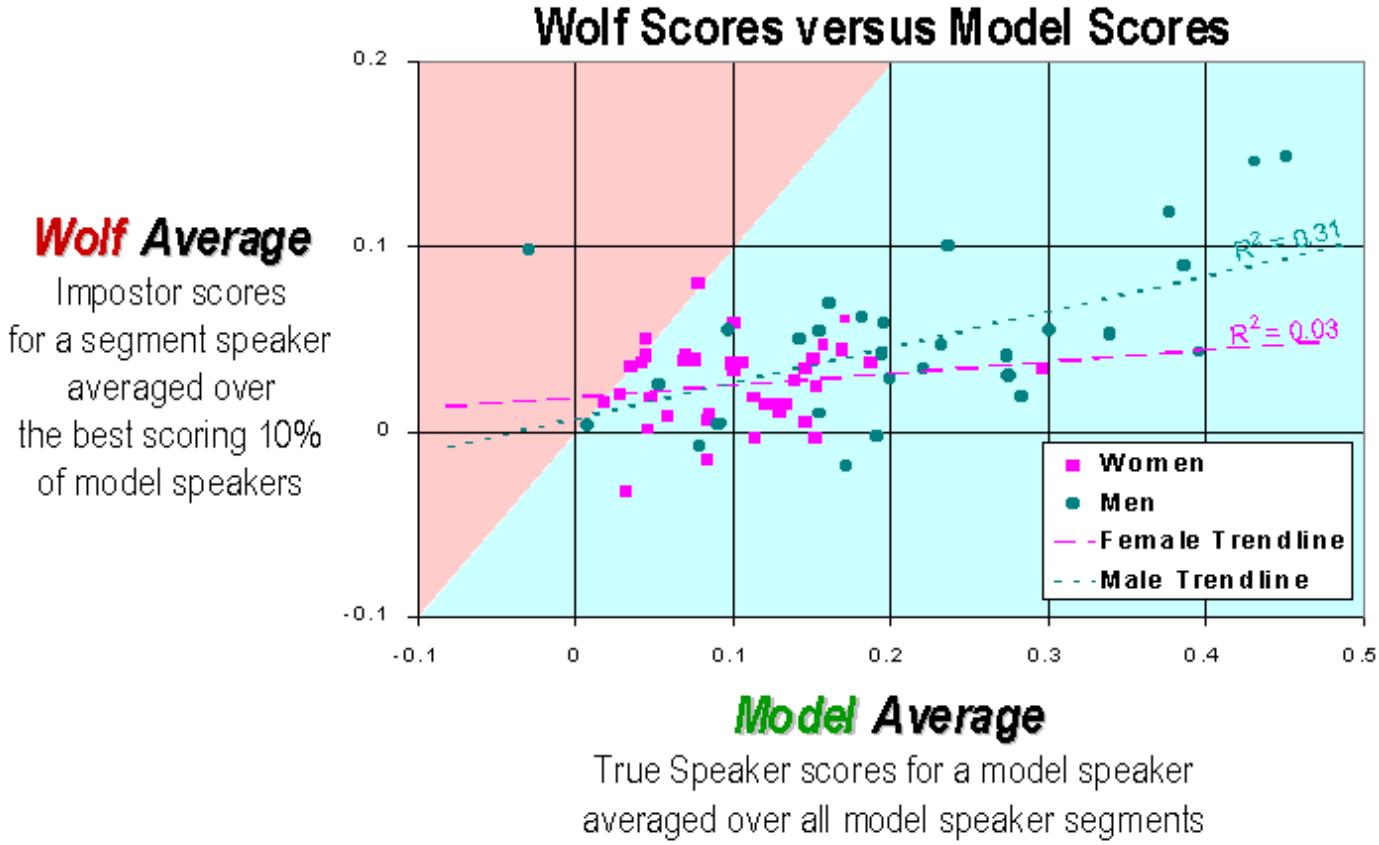
Misses for Model Speakers False Alarms for Model Speakers False Alarms for Segment Speakers

Cumulative Trials

ordered by Goat/Lamb/Wolf rank

Previous slide Next slide Back to first slide View graphic version

Can a Sheep be also a Wolf?

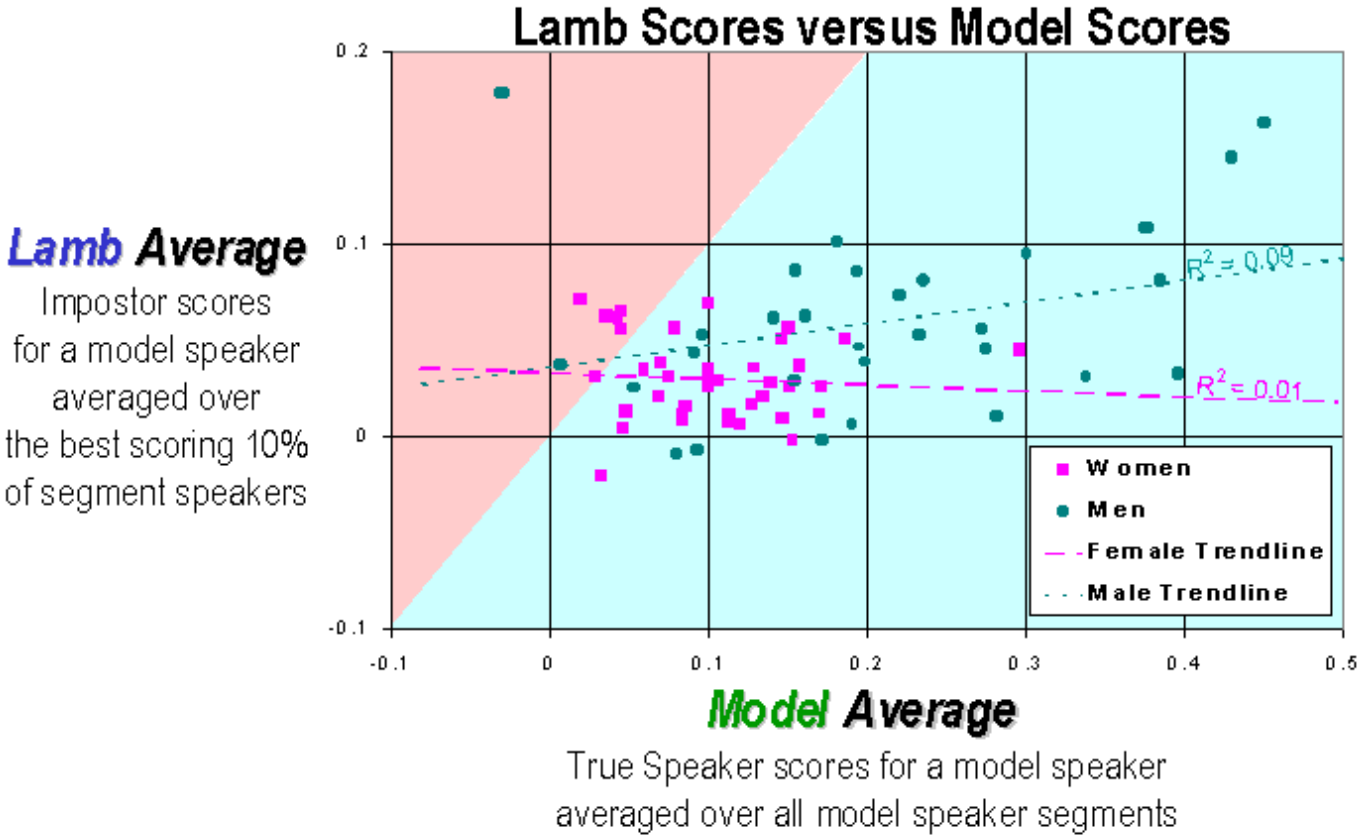


Can a Lamb be also a Wolf?

Wolf Scores versus Lamb Scores
Wolf Average
Impostor scores
for a segment speaker
averaged over
the best scoring 10%
of model speakers
Lamb Average
Impostor scores for a model speaker averaged over
the best scoring 10% of segment speakers
R2 = 0.63
R2 = 0.34

Previous slide Next slide Back to first slide View graphic version

Can a Sheep be also a Lamb?



Slide 13 of 14

Can a Sheep be also a Wolf?

Wolf Average

Impostor scores

for a segment speaker averaged over the best scoring 10% of model speakers

Model Average

True Speaker scores for a model speaker

averaged over all model speaker segments

Wolf Scores versus Model Scores

R2 = 0.31

R2 = 0.03

Previous slide Next slide Back to first slide View graphic version

Can a Sheep be also a Lamb?

Lamb Average

Impostor scores

for a model speaker averaged over the best scoring 10% of segment speakers

Model Average

True Speaker scores for a model speaker

averaged over all model speaker segments

Lamb Scores versus Model Scores

R2 = 0.01

R2 = 0.09

Previous slide Next slide Back to first slide View graphic version

Conclusion: It’s a out there!

Goats, Lambs and Wolves exist!

Recommendation:

- **Direct research toward the minimization of these undesirable animals.**
- **Include measurement of speaker differences as part of performance evaluation.**

An html copy of this presentation is available at URL http://www.nist.gov/speech/icslp_98

Previous slide Back to first slide View graphic version