浙江大学
ZHEJIANG UNIVERSITY

"OK, Siri" or "Hey, Google": Evaluating Voiceprint Distinctiveness Via Content-based *PROLE Score*

Ruiwen He, Xiaoyu Ji , Xinfeng Li, Yushi Cheng, and Wenyuan Xu
Zhejiang University, Tsinghua University

USENIX Security 2022

智能系统安全实验室
UBIQUITOUS SYSTEM SECURITY LAB.

Agenda

- Motivation
- Challenge and Methodology
- PROLE Score
- Finding and Conclusion

Motivation

- Voiceprint
  - Measurable characteristic of human voices
  - Popular biometric verification method

Unlock Phone

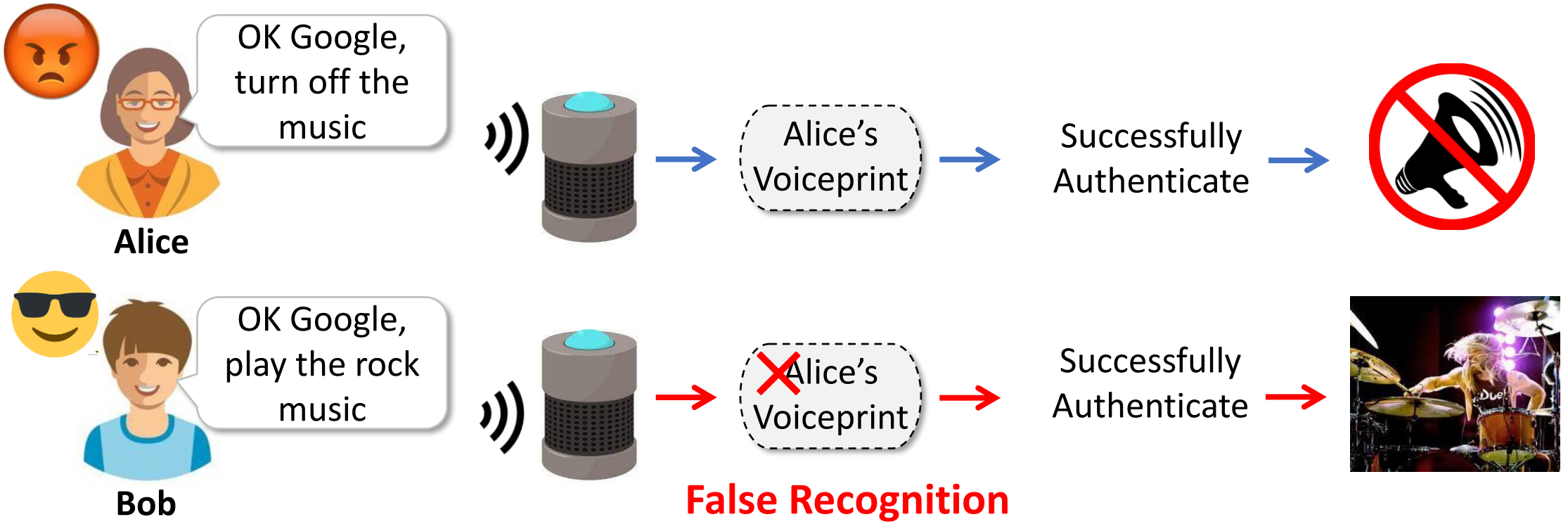
Authorize Payment

Wake up Smart Speaker

The security of voiceprint affects various industries

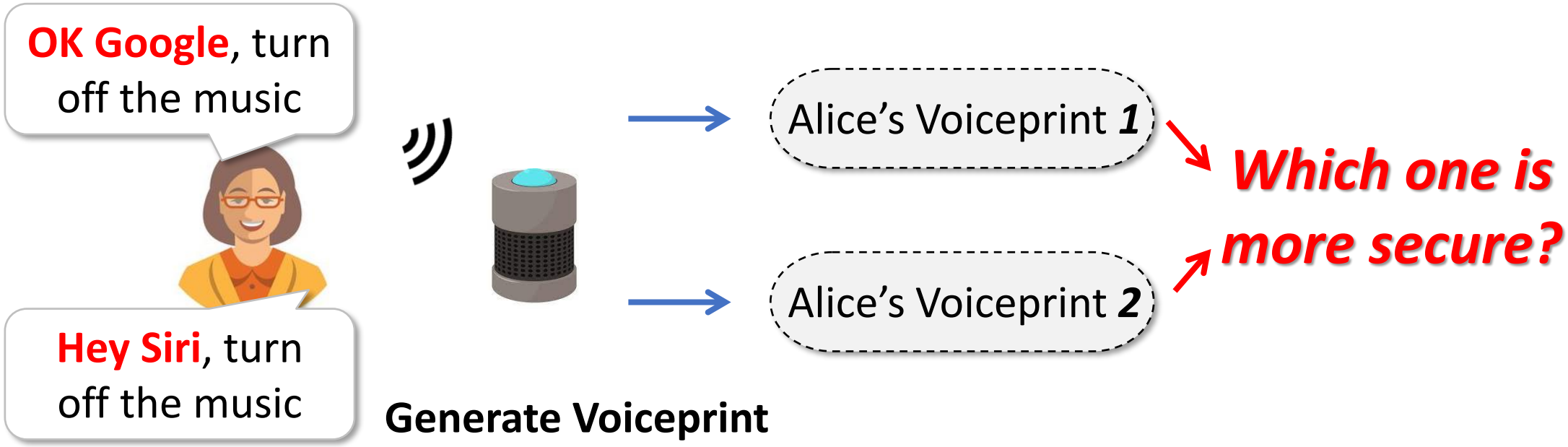
Security of Voiceprint

- Distinctiveness: Any two people don't have the same voiceprint



Security of Voiceprint

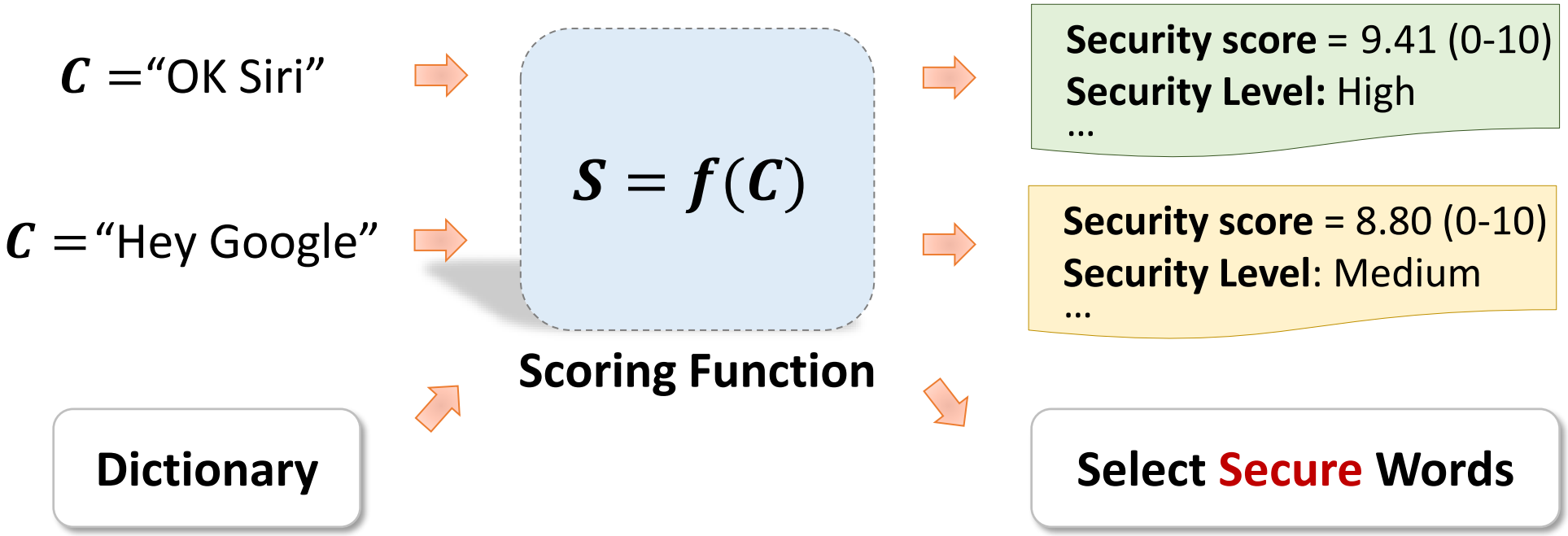
- Extracted voiceprint depends on the speech content



What types of speech content can provide secure voiceprints?

Goal

- Construct Voiceprint Security Scoring Metric
- Design more secure speech content for generating voiceprint



Goal

C ="OK Siri"

C ="Hey Google"

$S = f(C)$

Scoring Function

Security score = 9.41 (0-10)
Security Level: High
...

Security score = 8.80 (0-10)
Security Level: Medium
...

High Security Score → Statistically related to → **High** Distinctiveness → **Low** False Recognition Rate

- False Acceptance Rate (FAR)
- False Rejection Rate (FRR)

Typical Scenario

- Commercial wake-up words

OK Google | Alexa | Hey Siri | Hey Cortana | Xiaodu Xiaodu | Nihao XiaoE | .....

One or two words

With prefix word

Do they provide enough distinctiveness in voiceprints?

Manufacturer

Apply

Voiceprint Security Scoring Metric

Answer

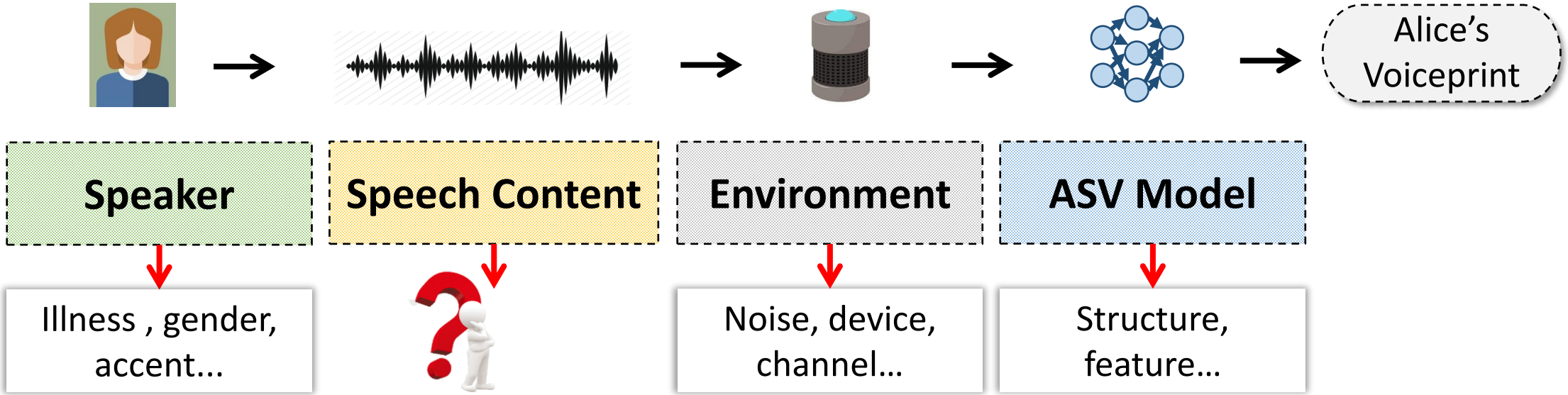
1. **Whether** wake-up words secure
2. **Why** they are secure/insecure
3. **How to** design secure words

Agenda

- Motivation
- Challenge and Methodology
- PROLE Score
- Finding and Conclusion

Challenge

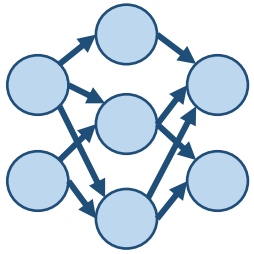
- Automatic Speaker Verification (ASV) System generate voiceprint
- Its **4 modules** affect the security of voiceprint



How to design a content-relevant scoring metric?

Challenge – Universal Scoring Metric

- *Challenge 1. Adapt to most models, speakers and environment*



- *Challenge 2. Handle a wide variety of speech content*

Basic idea – Statistical Analysis

- *Challenge 1. Adapt to most models, speakers and environment*

Solution

Large-scale Analysis: obtaining the statistical results of false recognition rate over a large dataset, including False Acceptance Rate (FAR) and False Rejection Rate (FRR)

3 Typical Models
- i-vector
- x-vector
- U-Level

2457 Speakers

Various genders, speech rates, channel noises..

1000-hour Utterances

2 Speech Datasets
- VCTK
- LibriSpeech

Basic idea – Statistical Analysis

- *Challenge 2. Handle a wide variety of speech content*

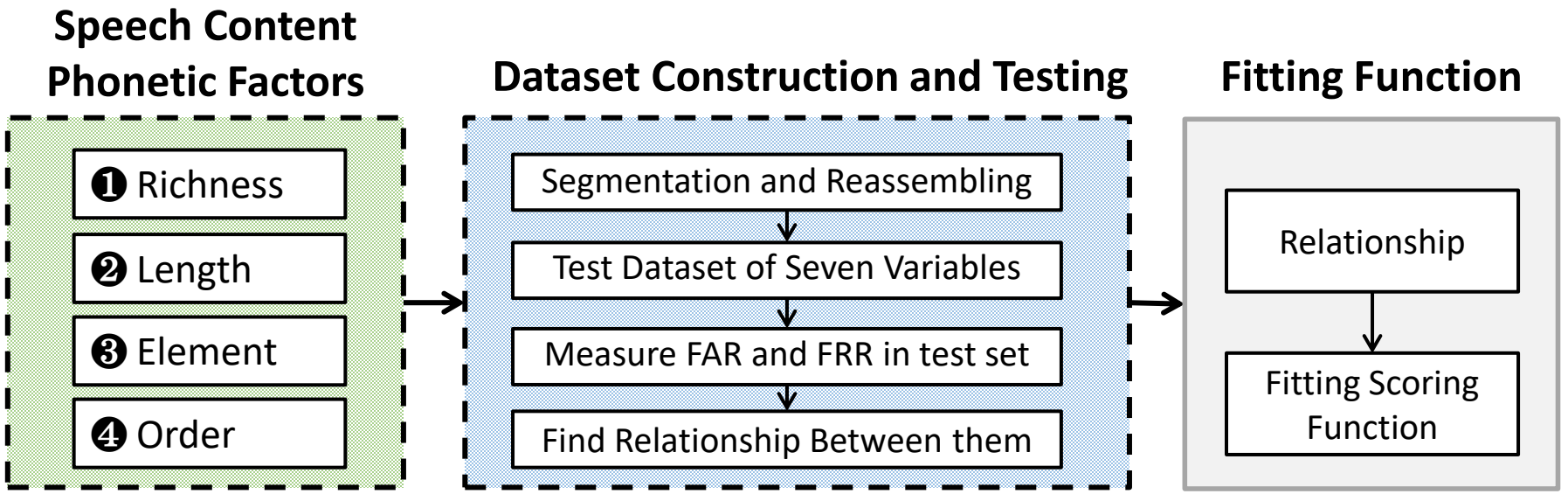
Solution

Comprehensive Speech Content: Composing arbitrary speech content by concatenating phones naturally

1000-hour Utterances → **Segment** to **Phone clips** → **Reassemble** to Test Audio → **14,600,000** audio samples

Measurement Methodology

- 3 Key Steps:
  - Speech Content Phonetic Factor Extracting
  - Dataset Construction and Testing
  - Function Fitting



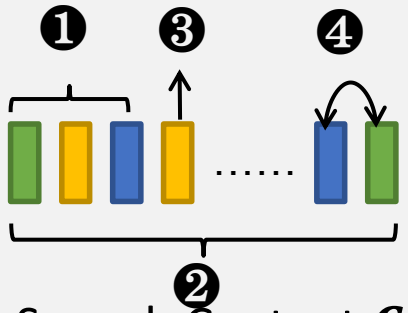
Agenda

- Motivation
- Challenge and Methodology
- PROLE Score
- Finding and Conclusion

PROLE Score – Voiceprint Security Metric

- Model the Speech Content
  - Basic unit of language – **Phone**, like [g], [k], [o]
  - Describe phones in words – **Phonetic Factors**

Phone Sequence



Phonetic Factors

❶ Richness R – Number of phone types

❷ Length L – Number of phones

❸ Element E – Specific type of phones

❹ Order O – sequential relationship among phones

Example: Hey Google

Phones : [heɪguːgəl]

R : 7

L : 8

E : [h][eɪ][g][u][ə][l]

O : [h] in the front of [eɪ]

PROLE Score – Voiceprint Security Metric

- Mapping the Phonetic Factors to Security Score
  - 4 variables: 4 Phonetic Factors
  - 3 implicit variables: Model, speaker, environment

Phonetic Factors

❶ Richness $\boldsymbol{R}$

❷ Length $\boldsymbol{L}$

❸ Element $\boldsymbol{E}$

❹ Order $\boldsymbol{O}$

$$\text{Score } \boldsymbol{S} = f(C) = f_{M,P,V}(R, L, E, O)$$

$\boldsymbol{M}$: Model, $\boldsymbol{P}$: Speaker, $\boldsymbol{V}$: Environment

PROLE Score – Voiceprint Security Metric

- Mapping the Phonetic Factors to Security Score
  - 4 variables: 4 Phonetic Factors
  - 3 implicit variables: Model, speaker, environment

Phonetic Factors

❶ Richness R

❷ Length L

❸ Element E

❹ Order O

PROLE Score Function

$$S = f(C) = f_{M,P,V}(R, L, E, O)$$

M: Model, P: Speaker, V: Environment

Parameterize PROLE Score

Phonetic Factors

❶ Richness R

❷ Length L

❸ Element E

❹ Order O

Statistical Results and Findings

Parameterize PROLE Score

Phonetic Factors

Statistical Results and Findings

❶ Repeat R

❷ Length L

❸ Encryption E

❹ Order O

1. Impact of Single Factor

2. Influence Ranking of 4 factors

Parameterize PROLE Score

Phonetic Factors

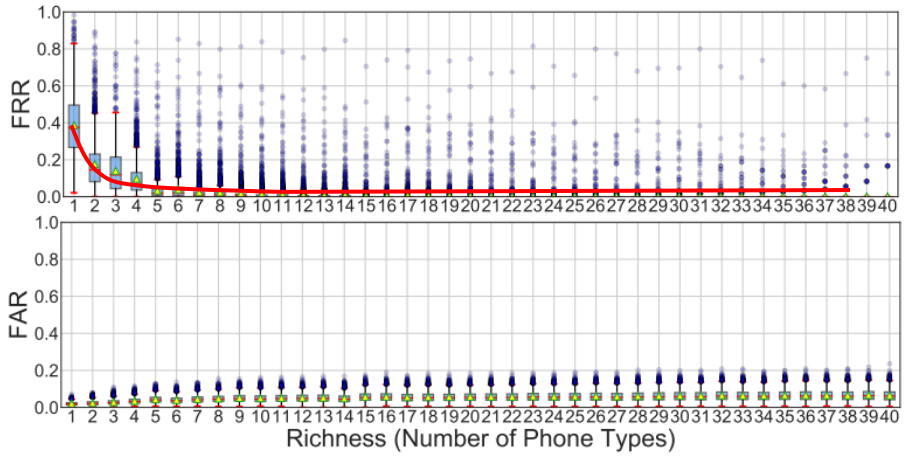
❶ Richness R

❷ Length L

❸ Element E

❹ Order O

Statistical Results and Findings



- As R increases, security increases
- $R > 5$, its impact is negligible
- Fitting function: $S(R) = ae^{bR} + ce^{dR}$

Finding1: Richness should be at least 5

Example

Content: Google

Phones : [guːgəl]

Richness: 4

Content: Hey Google

Phones : [heɪguːgəl]

Richness: 7

Parameterize PROLE Score

Phonetic Factors

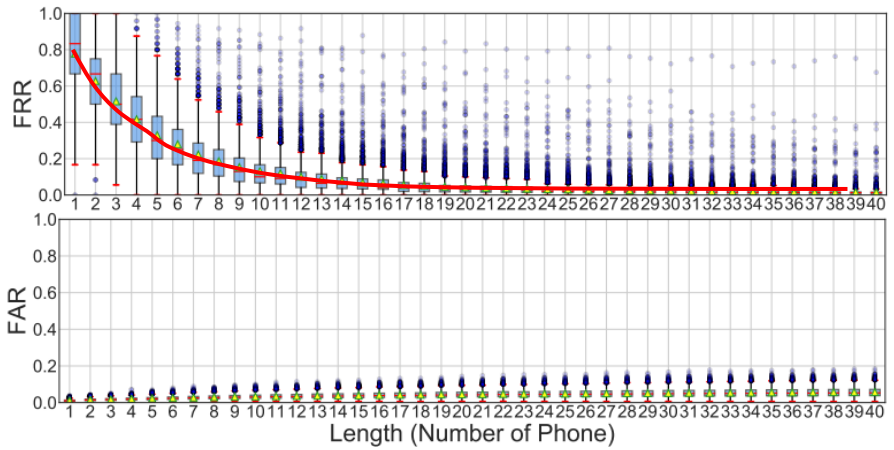
❶ Richness R

❷ Length L

❸ Element E

❹ Order O

Statistical Results and Findings



- As L increases, security increases
- L > 13, its impact is negligible
- Fitting function: $S(L) = ae^{bL} + c$

Finding2: Length should be at least 13

Example

Content: Xiaodu

Phones : [ɕjaʊtuɔ]

Length: 7

Content: Xiaodu **Xiaodu**

Phones : [ɕjaʊtuɔɕjaʊtuɔ]

Length: 14

Parameterize PROLE Score

Phonetic Factors

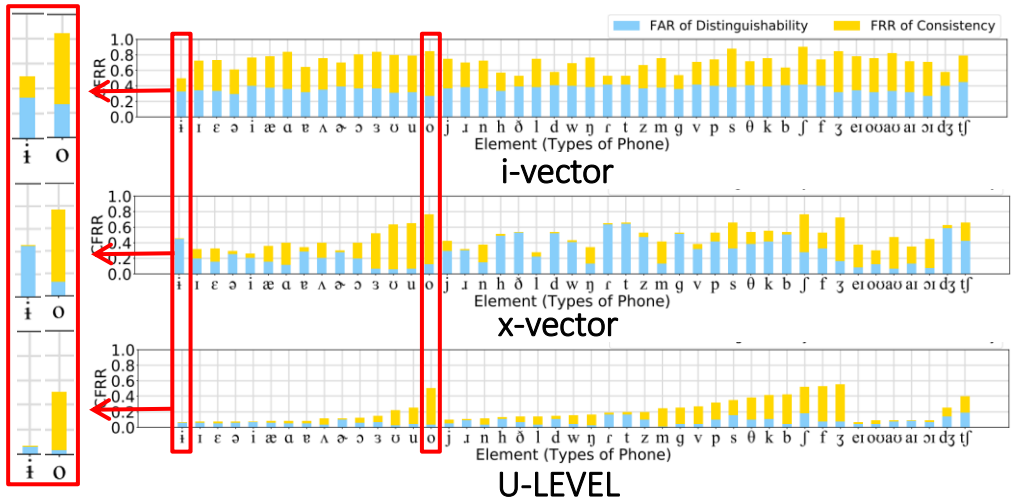
❶ Richness R

❷ Length L

❸ Element E

❹ Order O

Statistical Results and Findings



- Some phones are better for generating voiceprint

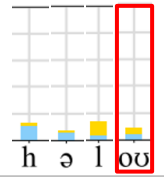
e.g., [ɨ] is better than [o] in 3 models

Finding3: We should choose secure phones

Example

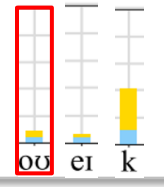
Content: Hello

Phones : [həloʊ]

Element: 

Content: OK

Phones : [oʊkeɪ]

Element: 

Parameterize PROLE Score

Phonetic Factors

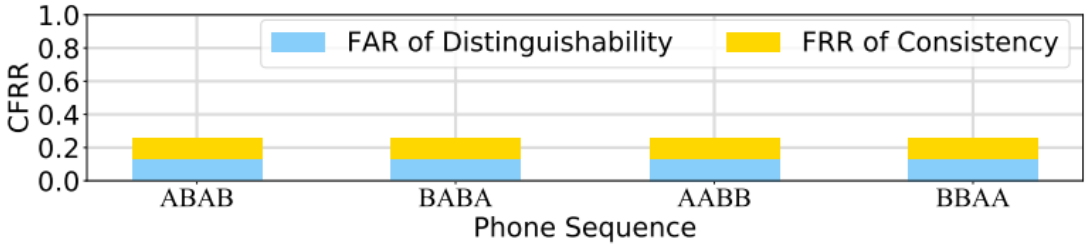
❶ Richness R

❷ Length L

❸ Element E

❹ Order O →

Statistical Results and Findings



- O has almost no effect on security

Note: A: secure phone, B: insecure phone

Finding4: We don't consider the order

1. Impact of Single Factor

2. Ranking of 4 factors

Richness > Element > Length>Order

Parameterize PROLE Score

- Fitting PROLE Scoring function with experimental data

$$\boldsymbol{S} = f(C)$$
$$= f_{M,P,V}(R, L, E, O)$$

$$S = F^{M}_{L+R}(R, L) + \frac{\sum_{\gamma}^{E} F^{M}_{E_\gamma}(L) L(\alpha)}{L} - F^{M}_{E_\alpha}(L)$$

Agenda

- Motivation
- Challenge and Methodology
- PROLE Score
- Finding and Conclusion

How to use the PROLE Score?

Example Scenario:
Select Wake-up words

Assessing Wake-up Words

- We assess 30 wake-up words from commercial products via PROLE Score.

English Wake-up Words Scores					Chinese Wake-up Words Scores				
Developer	Wake-up Words	i-v[1]	x-v[2]	U-L[3]	Developer	Wake-up Words	i-v	x-v	U-L
Amazon	Alexa	3.28	2.89	8.16	Alibaba	TianMaoJingLing	6.80	7.32	9.60
Amazon	Amazon	3.29	2.89	8.39	Baidu	XiaoDuXiaoDu	5.30	5.52	9.60
Amazon	Computer	4.61	4.38	8.68	Huawei	NiHaoXiaoE	4.90	5.35	9.93
Amazon	Echo	1.50	1.26	6.06	Huawei	NiHaoYoYo	4.72	4.63	9.27
Apple	Hey Siri	4.20	4.43	9.32	Huawei	XiaoEXiaoE	4.76	4.81	9.92
Google	Hey Google	4.33	4.56	8.88	JD	DingDongDingDong	4.61	4.38	8.55
Google	Ok Google	5.55	6.04	9.00	JD	Hey XiaoJingYu	6.86	7.49	9.82
Huawei	Hey Celia	4.44	4.21	9.24	Lenovo	NiHaoLianXiang	6.42	7.17	9.82
Microsoft	Hey Cortana	5.60	6.00	9.42	MeiZu	NiHaoMeiZu	6.06	6.25	9.52
Multiverse	Extreme	4.02	3.72	8.76	Microsoft	NiHaoXiaoNa	6.26	6.81	9.83
MyCroft	Hey Mycroft	5.57	6.01	9.02	Mobvoi	NiHaoWenWen	5.01	5.44	9.58
Nuance	Hello Dragon	6.01	6.76	9.46	OPPO	XiaoBuXiaoBu	5.17	5.39	9.43
OPPO	Hey Breeno	5.19	5.46	9.20	OPPO	XiaoOuXiaoOu	4.62	4.57	9.75
Samsung	Hey Bixby	4.85	4.79	9.05	Tencent	XiaoWeiXiaoWei	5.84	6.66	9.90
SoundHound	OK Hound	5.41	5.80	9.31	XiaoMi	XiaoAiTongXue	6.85	7.36	9.87

[1] Abbreviation for i-vector model.
[2] Abbreviation for x-vector model.
[3] Abbreviation for U-LEVEL model.

Findings – Commercial wake-up words

- 1. Enough richness

Wake-up word	Phone Sequence	Richness
Alexa	ɐlɛksə	6
Hey Bixby	heɪbɪksbi	7
Hey Breeno	heɪbɹiːnoʊ	9
Hey Celia	heɪsɛliə	8
Hey Cortana	heɪkɔːɹtɑːnə	10
OK Google	oʊkeɪguːgəl	9
Hey Siri	heɪsiəɹi	7
…		

>

Factor	Security Threshold
Richness	5

Findings – Commercial wake-up words

- 2. Inadequate length – Suggest to repeat

Wake-up word	Phone Sequence	Richness	Length
Alexa	ɐlɛksə	6	**6**
Hey Bixby	heɪbɪksbi	7	**9**
Hey Breeno	heɪbɹiːnoʊ	9	**9**
Hey Celia	heɪsɛliə	8	**8**
Hey Cortana	heɪkɔːɹtɑːnə	10	**10**
OK Google	oʊkeɪguːgəl	9	**10**
Hey Siri	heɪsiəɹi	7	**8**
…			

<

Factor	Security Threshold
Length	**13**

Findings – Commercial wake-up words

- 2. Inadequate length – Suggest to repeat

Wake-up words	Length	Score
Alexa	6	8.16

Repeat

Wake-up words	Length	Score
Alexa Alexa	12	9.15

Findings – Commercial wake-up words

- 3. Repeating in Chinese

English Wake-up Words Scores					Chinese Wake-up Words Scores				
Developer	Wake-up Words	i-v[1]	x-v[2]	U-L[3]	Developer	Wake-up Words	i-v	x-v	U-L
Amazon	Alexa	3.28	2.89	8.16	Alibaba	TianMaoJingLing	6.80	7.32	9.60
Amazon	Amazon	3.29	2.89	8.39	Baidu	XiaoDuXiaoDu	5.30	5.52	9.60
Amazon	Computer	4.61	4.38	8.68	Huawei	NiHaoXiaoE	4.90	5.35	9.93
Amazon	Echo	1.50	1.26	6.06	Huawei	NiHaoYoYo	4.72	4.63	9.27
Apple	Hey Siri	4.20	4.43	9.32	Huawei	XiaoEXiaoE	4.76	4.81	9.92
Google	Hey Google	4.33	4.56	8.88	JD	DingDongDingDong	4.61	4.38	8.55
Google	Ok Google	5.55	6.04	9.00	JD	Hey XiaoJingYu	6.86	7.49	9.82
Huawei	Hey Celia	4.44	4.21	9.24	Lenovo	NiHaoLianXiang	6.42	7.17	9.82
Microsoft	Hey Cortana	5.60	6.00	9.42	MeiZu	NiHaoMeiZu	6.06	6.25	9.52
Multiverse	Extreme	4.02	3.72	8.76	Microsoft	NiHaoXiaoNa	6.26	6.81	9.83
MyCroft	Hey Mycroft	5.57	6.01	9.02	Mobvoi	NiHaoWenWen	5.01	5.44	9.58
Nuance	Hello Dragon	6.01	6.76	9.46	OPPO	XiaoBuXiaoBu	5.17	5.39	9.43
OPPO	Hey Breeno	5.19	5.46	9.20	OPPO	XiaoOuXiaoOu	4.62	4.57	9.75
Samsung	Hey Bixby	4.85	4.79	9.05	Tencent	XiaoWeiXiaoWei	5.84	6.66	9.90
SoundHound	OK Hound	5.41	5.80	9.31	XiaoMi	XiaoAiTongXue	6.85	7.36	9.87

[1] Abbreviation for i-vector model. [2] Abbreviation for x-vector model. [3] Abbreviation for U-LEVEL model.

Findings – Commercial wake-up words

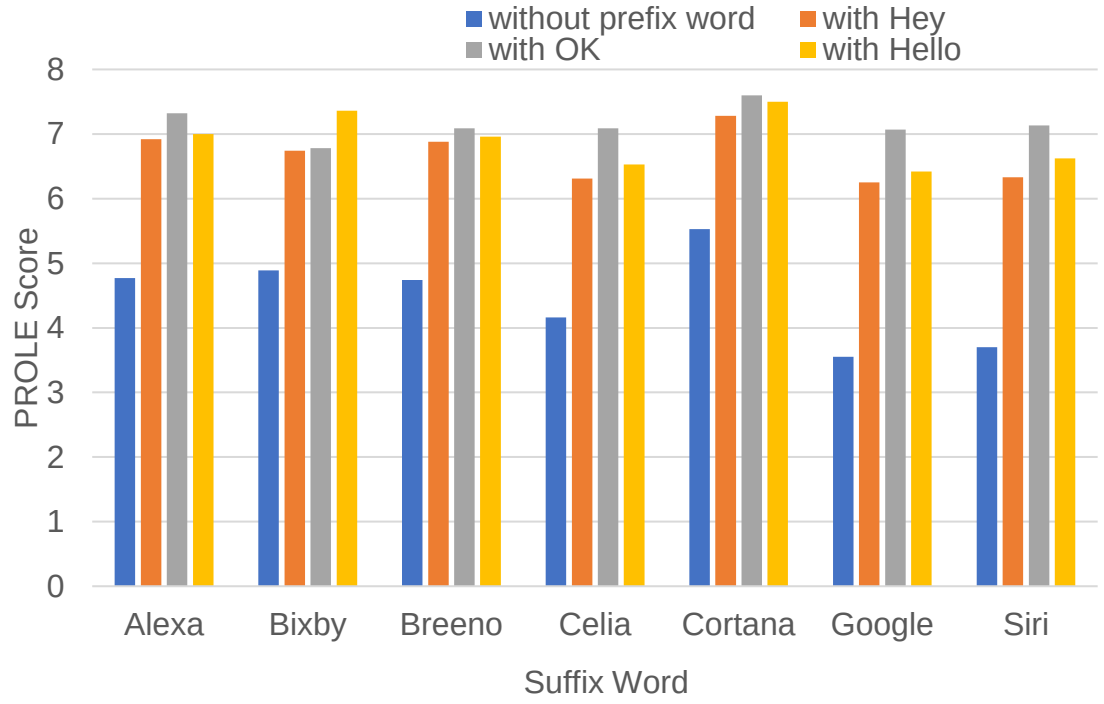
- 4. Prefix words in English – Adding magic phones

English Wake-up Words Scores					Chinese Wake-up Words Scores				
Developer	Wake-up Words	i-v[1]	x-v[2]	U-L[3]	Developer	Wake-up Words	i-v	x-v	U-L
Amazon	Alexa	3.28	2.89	8.16	Alibaba	TianMaoJingLing	6.80	7.32	9.60
Amazon	Amazon	3.29	2.89	8.39	Baidu	XiaoDuXiaoDu	5.30	5.52	9.60
Amazon	Computer	4.61	4.38	8.68	Huawei	NiHaoXiaoE	4.90	5.35	9.93
Amazon	Echo	1.50	1.26	6.06	Huawei	NiHaoYoYo	4.72	4.63	9.27
Apple	Hey Siri	4.20	4.43	9.32	Huawei	XiaoEXiaoE	4.76	4.81	9.92
Google	Hey Google	4.33	4.56	8.88	JD	DingDongDingDong	4.61	4.38	8.55
Google	Ok Google	5.55	6.04	9.00	JD	Hey XiaoJingYu	6.86	7.49	9.82
Huawei	Hey Celia	4.44	4.21	9.24	Lenovo	NiHaoLianXiang	6.42	7.17	9.82
Microsoft	Hey Cortana	5.60	6.00	9.42	MeiZu	NiHaoMeiZu	6.06	6.25	9.52
Multiverse	Extreme	4.02	3.72	8.76	Microsoft	NiHaoXiaoNa	6.26	6.81	9.83
MyCroft	Hey Mycroft	5.57	6.01	9.02	Mobvoi	NiHaoWenWen	5.01	5.44	9.58
Nuance	Hello Dragon	6.01	6.76	9.46	OPPO	XiaoBuXiaoBu	5.17	5.39	9.43
OPPO	Hey Breeno	5.19	5.46	9.20	OPPO	XiaoOuXiaoOu	4.62	4.57	9.75
Samsung	Hey Bixby	4.85	4.79	9.05	Tencent	XiaoWeiXiaoWei	5.84	6.66	9.90
SoundHound	OK Hound	5.41	5.80	9.31	XiaoMi	XiaoAiTongXue	6.85	7.36	9.87

[1] Abbreviation for i-vector model.
[2] Abbreviation for x-vector model.
[3] Abbreviation for U-LEVEL model.

Findings – Commercial wake-up words

- 4. Prefix words in English – Adding magic phones
  - We compare the average PROLE scores of 7 wake-up words with and without the prefix words.

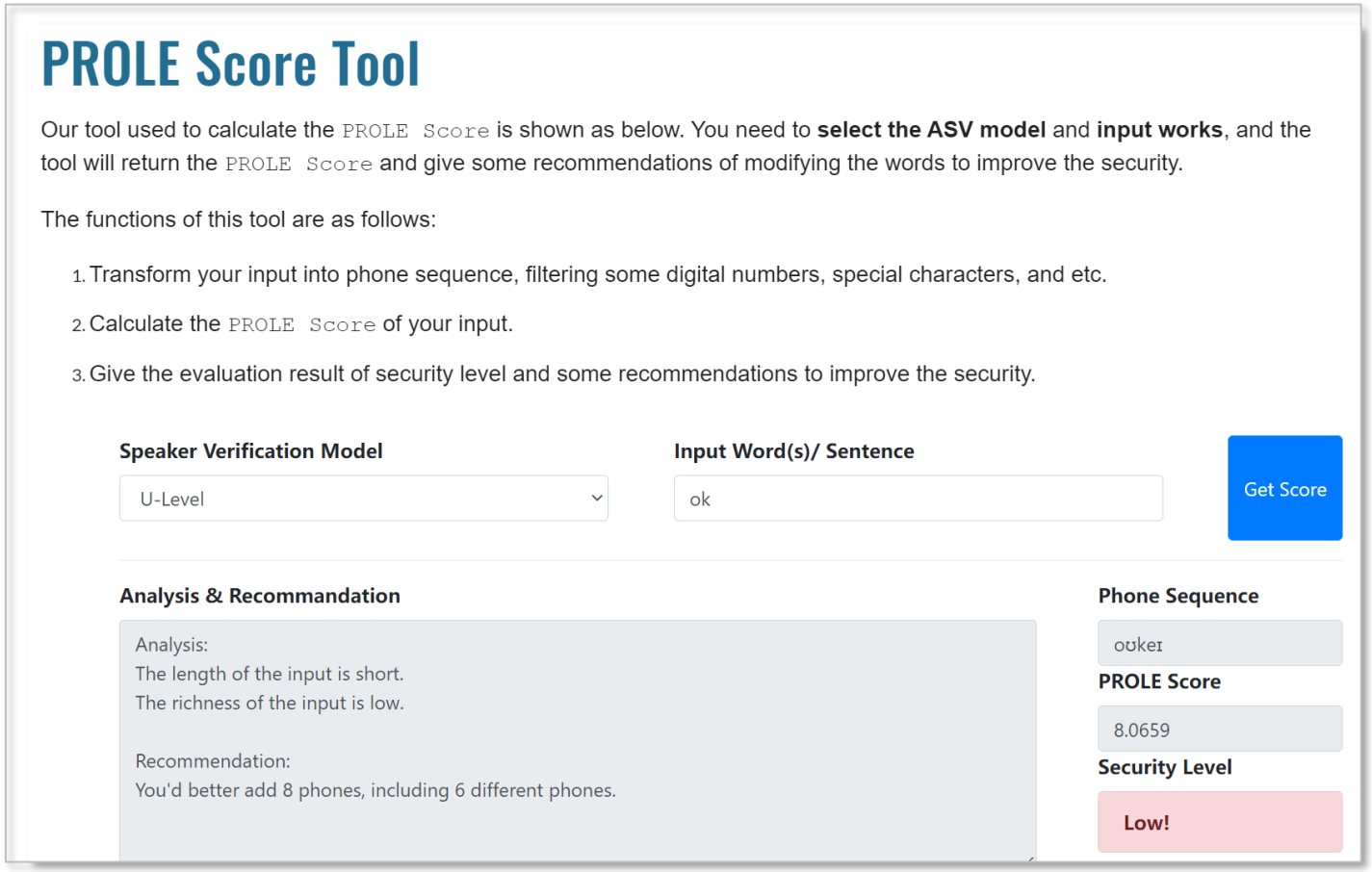


Prefix word	Phone Sequence	Magic Phones
Hey	heɪ	[h][eɪ]
OK	oʊkeɪ	[oʊ][eɪ]
Hello	həloʊ	[h][ə][oʊ]

- Increase PROLE Score by around 24%
- Recommended prefix word: OK

A Universal Scoring Tool

- We develop an online voiceprint security scoring tool.
  - ***Input:*** Identification words, model
  - ***Output:*** A security score, suggestion
  - https://sites.google.com/view/voiceprint-sec



PROLE Score Tool

Our tool used to calculate the `PROLE Score` is shown as below. You need to **select the ASV model** and **input works**, and the tool will return the `PROLE Score` and give some recommendations of modifying the words to improve the security.

The functions of this tool are as follows:

1. Transform your input into phone sequence, filtering some digital numbers, special characters, and etc.
2. Calculate the `PROLE Score` of your input.
3. Give the evaluation result of security level and some recommendations to improve the security.

Speaker Verification Model

U-Level

Input Word(s)/ Sentence

responsibility

Get Score

Analysis & Recommandation

Phone Sequence

ɹɪspɑːnsəbɪlɪɹi

PROLE Score

9.4735

Security Level

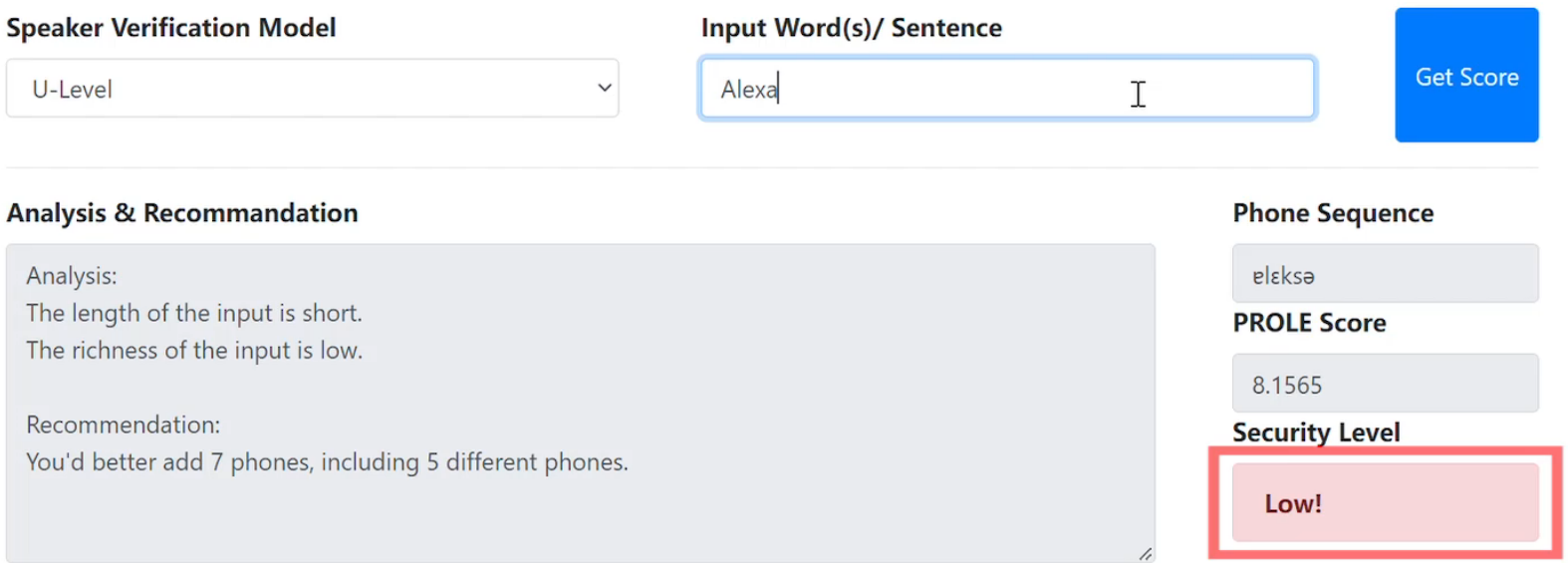
High!

PROLE Score Tool

Our tool used to calculate the `PROLE Score` is shown as below. You need to **select the ASV model** and **input works**, and the tool will return the `PROLE Score` and give some recommendations of modifying the words to improve the security.

The functions of this tool are as follows:

1. Transform your input into phone sequence, filtering some digital numbers, special characters, and etc.
2. Calculate the `PROLE Score` of your input.
3. Give the evaluation result of security level and some recommendations to improve the security.

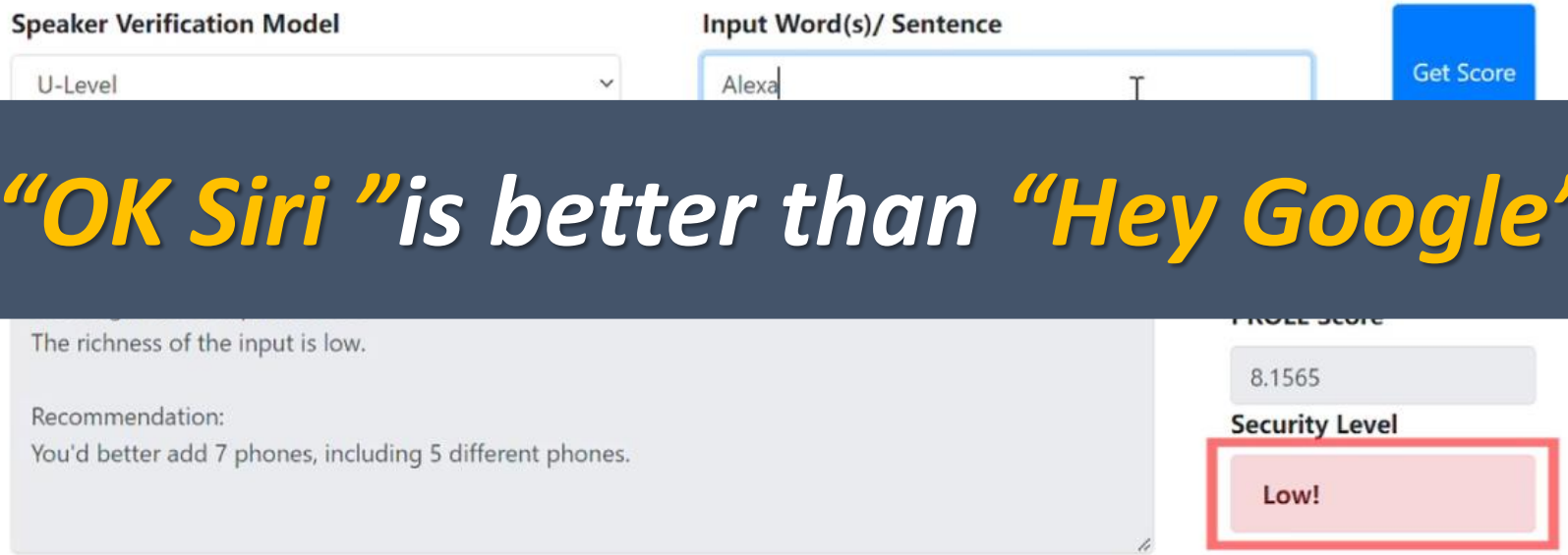


PROLE Score Tool

Our tool used to calculate the PROLE Score is shown as below. You need to **select the ASV model** and **input works**, and the tool will return the PROLE Score and give some recommendations of modifying the words to improve the security.

The functions of this tool are as follows:

1. Transform your input into phone sequence, filtering some digital numbers, special characters, and etc.
2. Calculate the PROLE Score of your input.
3. Give the evaluation result of security level and some recommendations to improve the security.



"OK Siri "is better than "Hey Google"

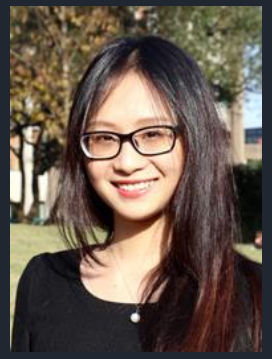
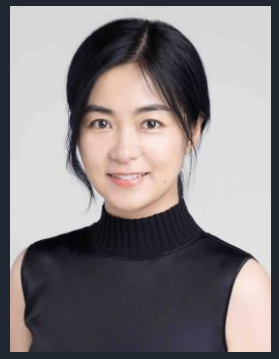
Conclusions

- Explored the correlation between the security of voiceprint and the speech content.
- Proposed a scoring metric called PROLE Score to evaluate the secure of voiceprint.
- Manufacturers and users can use it to select secure wake-up words.

- Future plan:
  - Study the effect of accent
  - Improve the transferability of metric across models

浙江大学 ZHEJIANG UNIVERSITY
智能系统安全实验室 UBIQUITOUS SYSTEM SECURITY LAB.

"OK, Siri" or "Hey, Google": Evaluating Voiceprint Distinctiveness Via Content-based *PROLE Score*

USSLAB Website: http://usslab.org/

Contact authors:

wyxu@zju.edu.cn

xji@zju.edu.cn

BACK UP

Parameterize PROLE Score

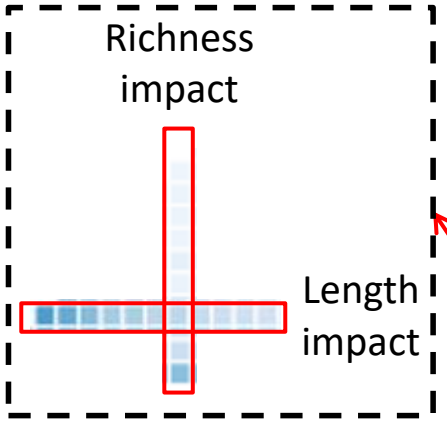
Phonetic Factors

❶ Richness R

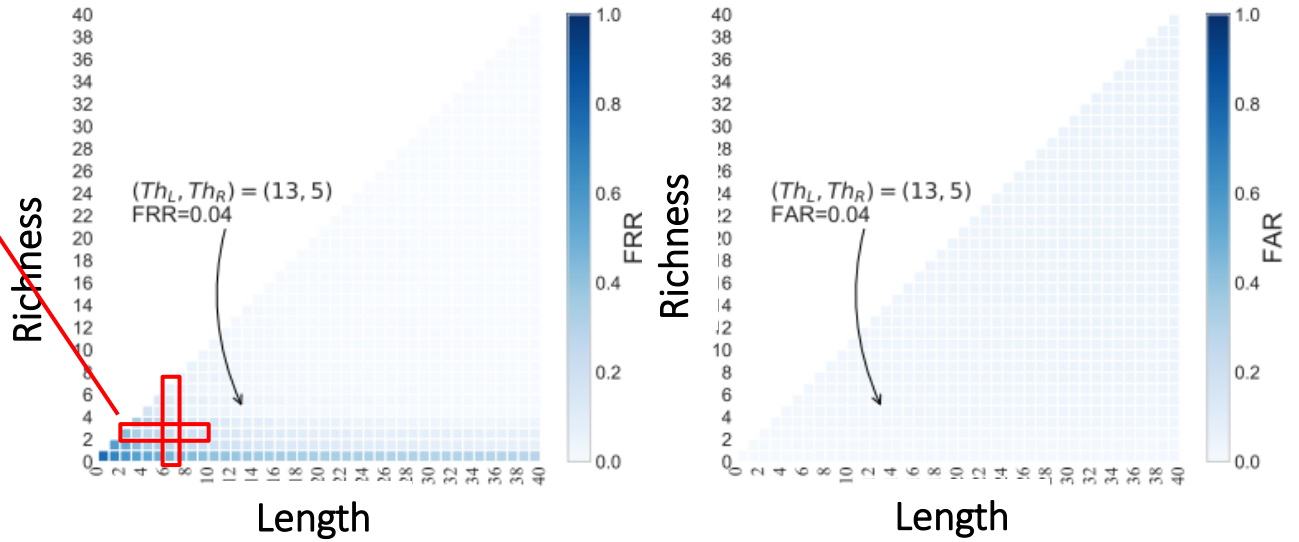
❷ Length L

❸ Element E

❹ Order O



Statistical Results and Findings



- Gradient in R is higher than it in L
- Fitting function: $S(R, L) = ae^{bL} + ce^{dR} + m$

Finding5: Richness > length

Parameterize PROLE Score

Phonetic Factors

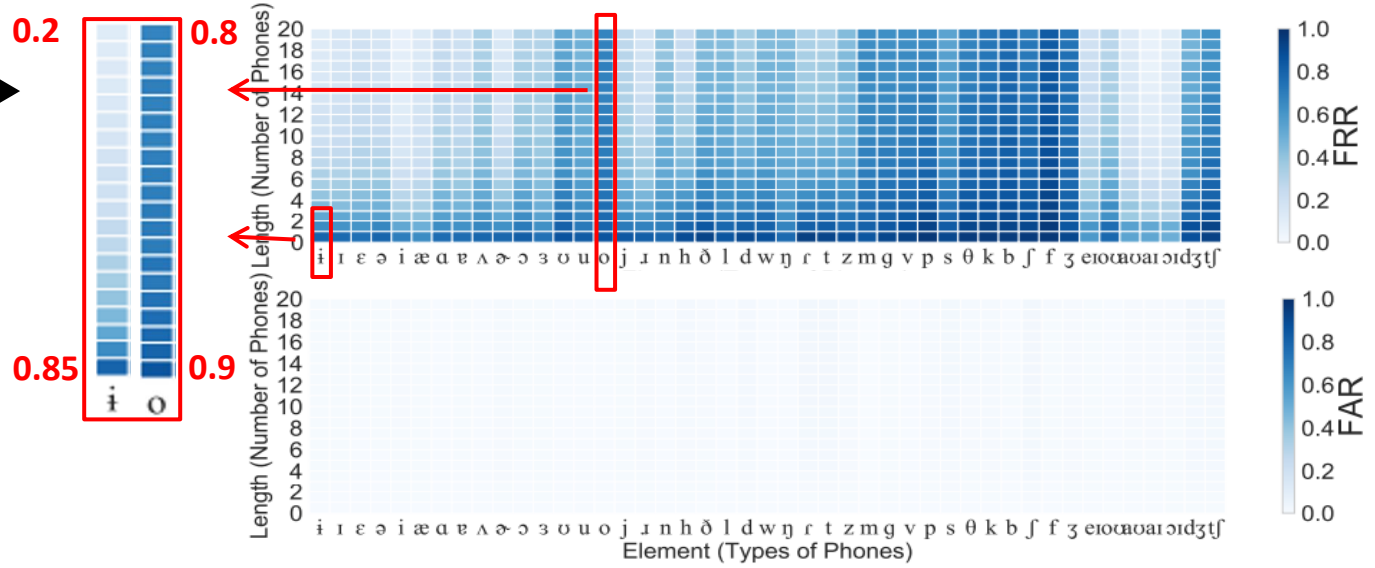
❶ Richness R

❷ Length L

❸ Element E

❹ Order O

Statistical Results and Findings



- Repeat insecure phone → Little improvement
- Fitting function: $S_E(L) = ae^{bL} + c$

Finding6: Element > Length

Parameterize PROLE Score

Phonetic Factors

❶ Richness R

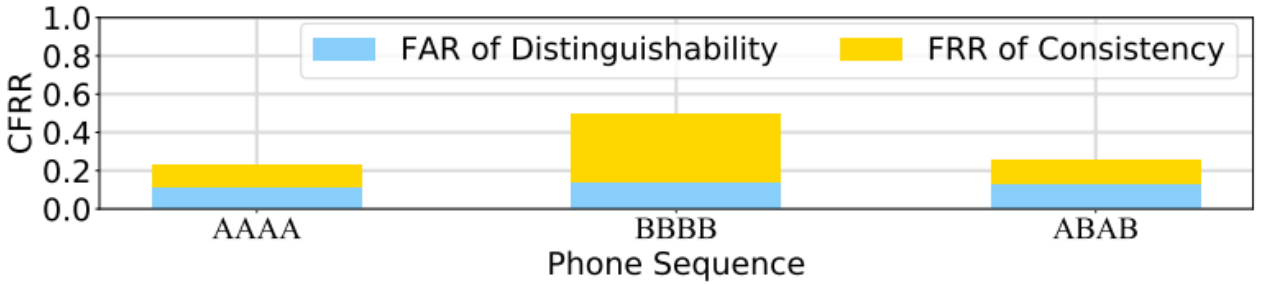
❷ Length L

❸ Element E

❹ Order O

→

Results and Findings



- Repeating the most secure phone "$AAAA$"
  = Adding the least secure phone "$ABAB$"
  < Adding an arbitrary phone "$AxAx$"

Note: A: secure phone, B: insecure phone

Finding7: Richness > Element

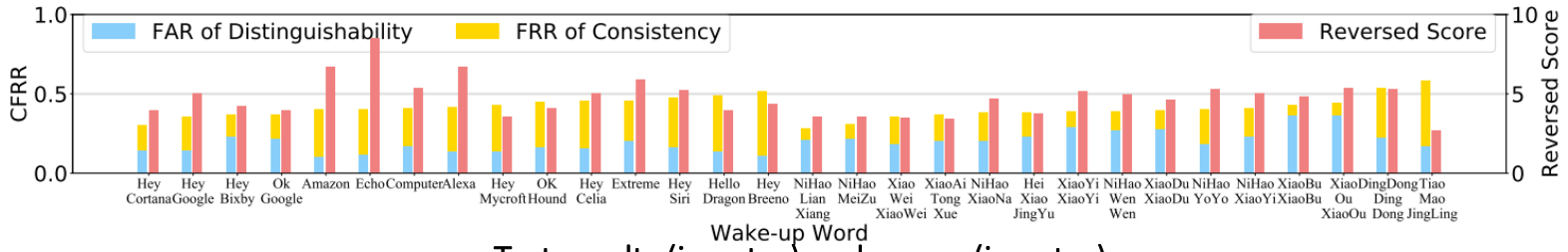
Effectiveness Validation of PROLE Score

- We validate the effectiveness of PROLE score via a user study.

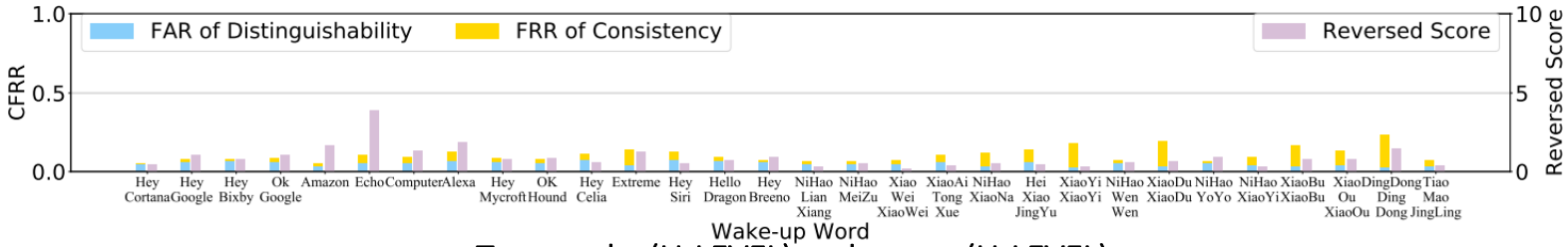
- Setup:
  - 46 volunteers
  - Aged between 19 and 50
  - 23 females and 23 males
  - Read wake-up words 3 times
  - Record with 4 smartphones

- Correlation:

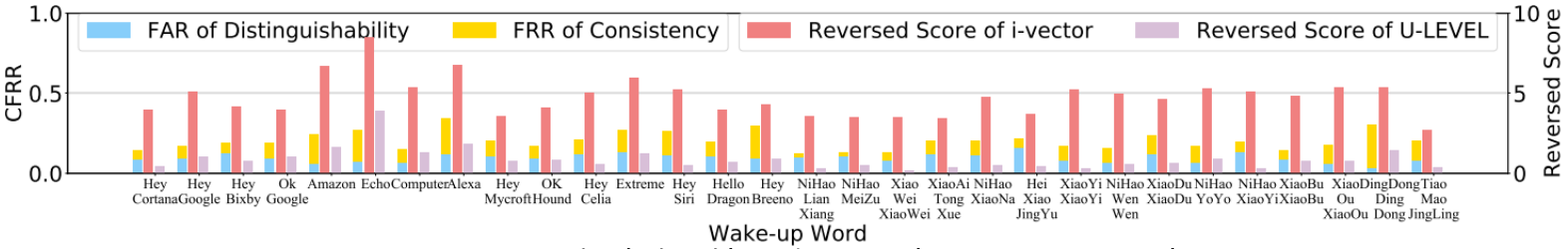
(distance correlation coefficient)
  - 0.781 for i-vector
  - 0.748 for U-LEVEL
  - >0.78 for iFlytek and score under 2 models



Test results (i-vector) and scores (i-vector)



Test results (U-LEVEL) and scores (U-LEVEL)



Test results (iFlytek) and scores (i-vector, U-LEVEL)

PROLE Score Tool

Our tool used to calculate the `PROLE Score` is shown as below. You need to **select the ASV model** and **input works**, and the tool will return the `PROLE Score` and give some recommendations of modifying the words to improve the security.

The functions of this tool are as follows:

1. Transform your input into phone sequence, filtering some digital numbers, special characters, and etc.
2. Calculate the `PROLE Score` of your input.
3. Give the evaluation result of security level and some recommendations to improve the security.

Speaker Verification Model

U-Level

Input Word(s)/ Sentence

e.g. OK Google

Get Score

Analysis & Recommandation

Phone Sequence

PROLE Score

Security Level

Unknown!