

Consonant-Vowel-Consonants for Error-Free Code Entry

Nikola K. Blanchard^{1,2}, Leila Gabasova³, Ted Selker⁴

¹Institut de Recherche en Informatique Fondamentale, Université Paris Diderot

²Digitrust, Loria, Université de Lorraine

³Institut de Plantologie et d'Astrophysique de Grenoble

⁴University of Maryland, Baltimore County

First International Conference on HCI for Cybersecurity, Privacy and Trust @ HCII

July 29th, 2019

Introduction: a voting experiment

Voting experiments in Strasbourg and San-Sebastian



Ballots at the Global Forum on Modern Direct Democracy

Random-Sample Voting Ballot

QUESTION: Should voting in national elections be compulsory?

VOTING TIME: 12:00PM CET Thursday 17 November 2016 through 9:30PM CET Friday 18 November 2016

INSTRUCTIONS:

- 1 Choose either half of this sheet randomly (ballot number and password are the same for both halves).
- 2 Use a web browser to visit the webpage: <https://vbb.rsvoting.org/rsv/vbb/gfadd2016-g1/>
Your ballot number is your **login** **⓪**: 001
Your **password** **⓪** is: vhbz-buhb-mrda-fwps
- 3 When prompted, enter the vote code that is printed adjacent your vote.
- 4 You should discard or destroy at least the half of this sheet that you used to vote; it is recommended, however, that you keep the other half of this sheet and write down on it in the space provided your vote code for later use in the audit.

Choice	Vote-Code ⓪
Yes	4457-1444-2131
No	6975-7435-2625



QUESTION: Should voting in national elections be compulsory?

VOTING TIME: 12:00PM CET Thursday 17 November 2016 through 9:30PM CET Friday 18 November 2016

INSTRUCTIONS:

- 1 Choose either half of this sheet randomly (ballot number and password are the same for both halves).
- 2 Use a web browser to visit the webpage: <https://vbb.rsvoting.org/rsv/vbb/gfadd2016-g1/>
Your ballot number is your **login** **⓪**: 001
Your **password** **⓪** is: vhbz-buhb-mrda-fwps
- 3 When prompted, enter the vote code that is printed adjacent your vote.
- 4 You should discard or destroy at least the half of this sheet that you used to vote; it is recommended, however, that you keep the other half of this sheet and write down on it in the space provided your vote code for later use in the audit.

Choice	Vote-Code ⓪
Yes	4134-9733-6914
No	1855-4750-4118

Random-Sample Voting Ballot

Random-Sample Voting Ballot

QUESTION: Should voting in national elections be compulsory?

VOTING TIME: 12:00PM CET Thursday 17 November 2016 through 9:30PM CET Friday 18 November 2016

INSTRUCTIONS:

- 1 Choose either half of this sheet randomly (ballot number and password are the same for both halves).
- 2 Use a web browser to visit the webpage: <https://vbb.rsvoting.org/rsv/vbb/gfmd2016-q1/>
Your ballot number is your **login ①**: 001
Your **password ②** is: **vhbe-buhb-mrda-fwpz**
- 3 When prompted, enter the vote code that is printed adjacent your vote.
- 4 You should discard or destroy at least the half of this sheet that you used to vote; it is recommended, however, that you keep the other half of this sheet and write down on it in the space provided your vote code for later use in the audit.

Choice	Vote-Code ③
Yes	4457-1444-2131
No	6975-7435-2625

Experiment design

5 sections with layers of A/B testing on order and content

- Welcome and basic information
- Transcription: 9 codes – 3 structures and 3 lengths
- Choice: 9 pairs of codes (alphanumeric vs alternative), choose one to transcribe
- Memory: 7 codes, users asked whether they'd seen it before
- Confirm and send data

Lengths from 9 to 22, with 4 main structures:

- *Numeric*: 958905239
- *Alphabetic*: lower-case Latin letters: ienkzeiwa
- *Alphanumeric*: numbers and mixed-case characters: Ok9Kh51ml
- *CVCs*: consonant-vowel-consonant alphabetic trigrams in lower-case: cofbujkilzaz

Welcome

Basic Info

Section 1

Section 2

Section 3

Transcription

Please manually retype the password from the left into the field on the right.

331 339 217

Next

Welcome

Basic Info

Section 1

Section 2

Section 3

Choice

Please choose one of the next two codes (the one that seems the easiest to write) and type it in the field underneath.

IBK 00CP q9n

leh wyl zyx zuj xic

Next

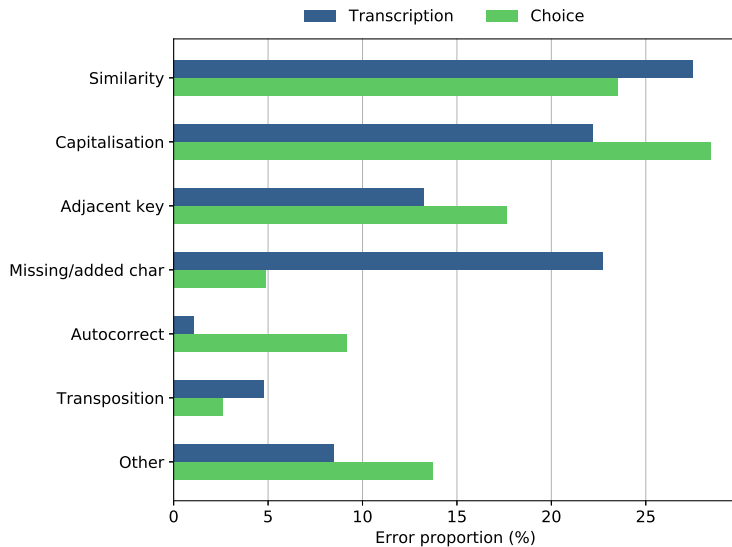
33 participants in pilot study, 267 participants in follow-up. 3 main groups (by recruitment method):

- 115 respondents from online psychology portal, overwhelmingly from USA
- 91 French in snowball sampling from tech networks
- 61 international from general social networks (24 countries, 14 languages)

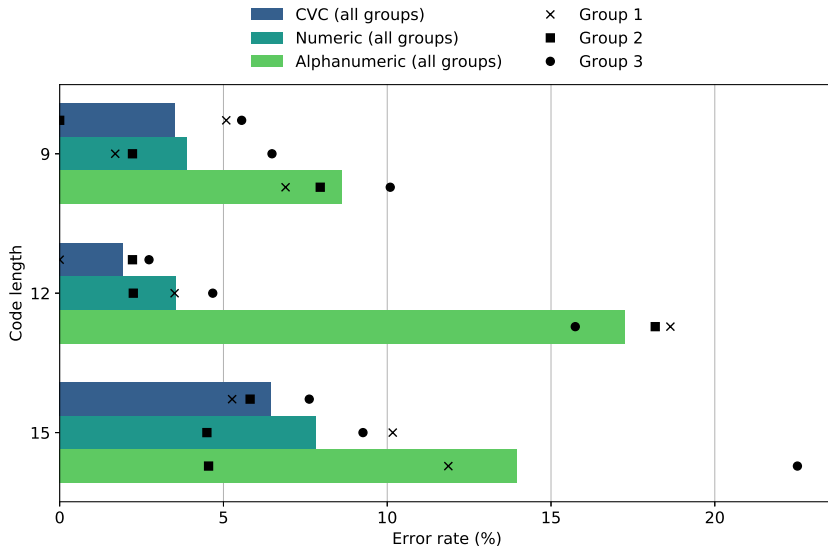
Multiple questions:

- How does structure and length affect error frequency?
- How does structure and length affect typing speed?
- How does structure and length affect memorability of the code?
- Are alphanumeric codes optimal for some metrics?
- What is the impact of chunking?

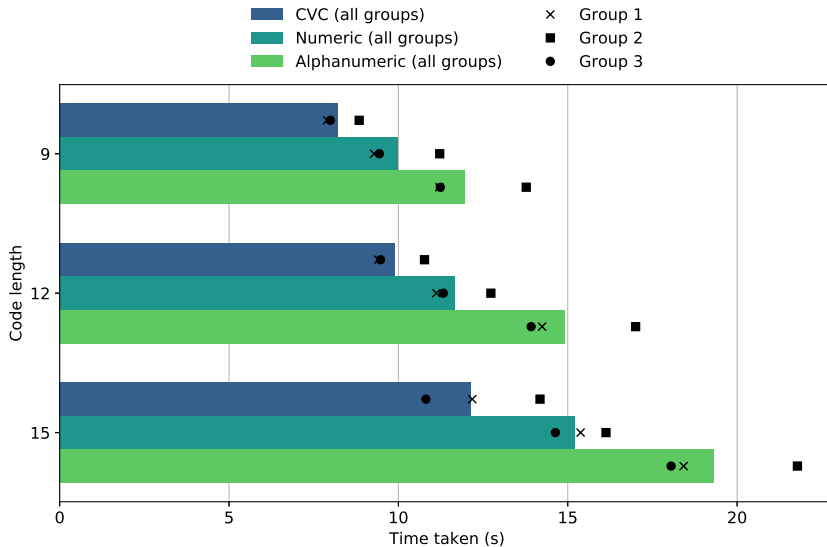
Error types



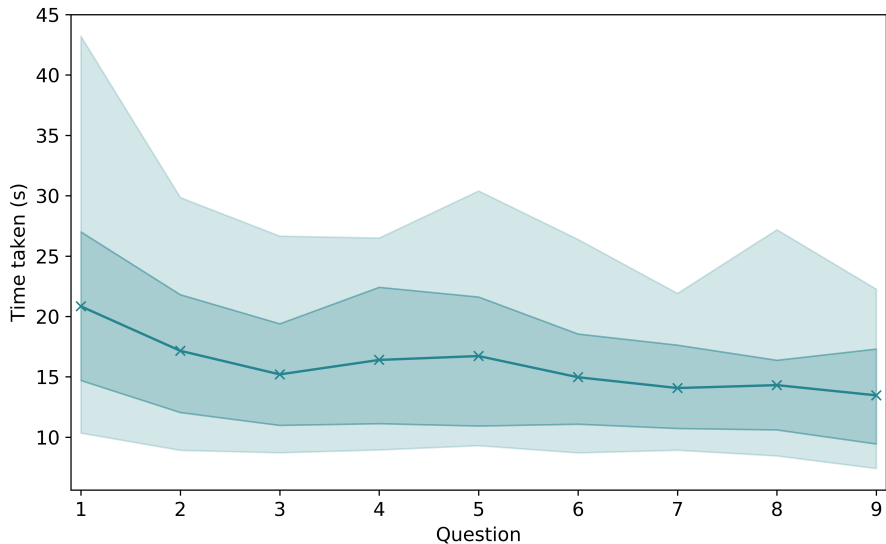
Transcription: error rates by structure and length



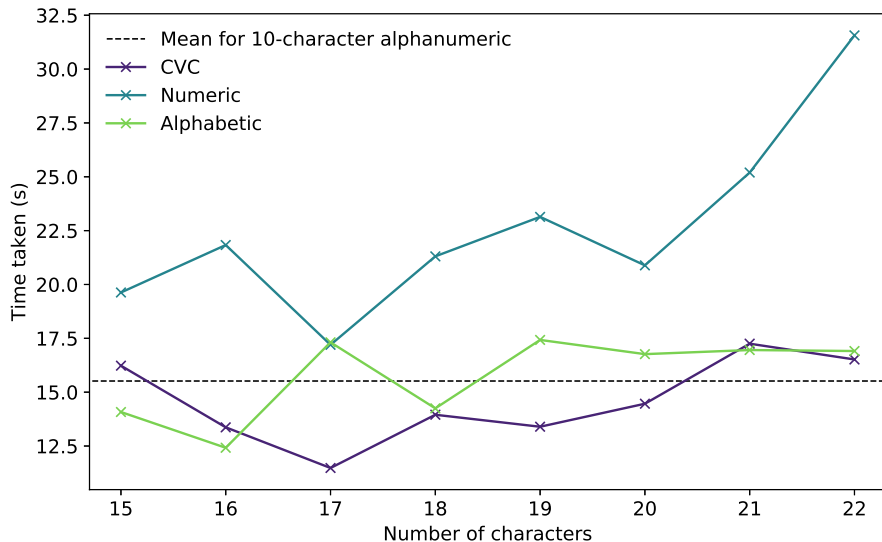
Transcription: speed by structure and length



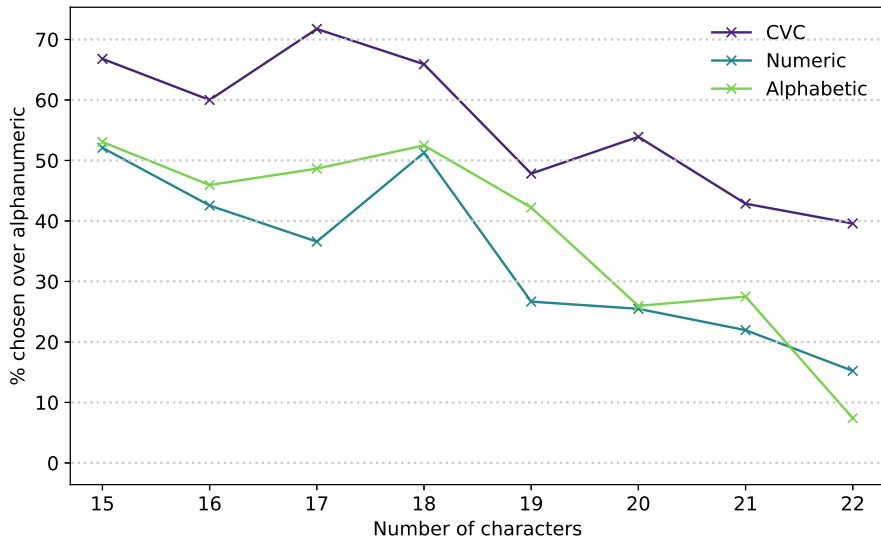
Choice: alphanumeric speed



Choice: alternative code speed



Code preference against alphanumeric



267 participants, 121 patterns, more than 35% of users choosing among these:

- 31 always chose the alphanumeric
- 24 chose the alphanumeric for all cases but one (either short or mid-length CVCs or numeric)
- 18 only chose the alphanumeric against numeric codes
- 12 only chose the alphanumeric in one case
- 11 never chose the alphanumeric

Proportion of errors recalling the code in the Memory section:

Error type	NUM9	CVC9	CVC12	CVC15	ANUM9	ANUM12	ANUM15
Type 1	28.6	39.0	6.5	19.0	40.1	18.4	25.7
Type 2	15.7	18.3	10.9	9.4	17.7	6.4	5.8
Total	22.5	28.8	8.6	14.4	29.2	12.0	16.7

25% of false positives, 13% of false negatives.

Making better codes:

Making a better code:

Lessons:

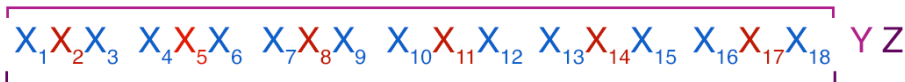
- use a fixed length to detect length errors
- avoid certain characters such as q or g
- avoid alphanumeric and capitalisation
- syllabic codes seem to have an advantage

CVCs seem to work well, one question is the length.

Error correction

consonant (except q): equivalent to 1-19

$$Y = \sum X_i \text{ mod } 19 \text{ (sum)}$$



vowel: equivalent to 1-6

$$Z = \sum i \cdot X_i \text{ mod } 19 \text{ (weighted sum)}$$

Advantages of CVC⁶ :

- More entropy than 10-character alphanumeric (66.5 vs 59.5 bits)
- Faster by more than 10%
- Preferred by at least 2/3 of users
- Normal errors below 5%
- Error correction can make it less than 0.2%

Contributions:

- First systematic study of structure effect on transcription error and speed
- Alphanumeric codes are bad on most metrics
- The trade-off for syllabic codes is worth the length
- CVC offers a good alternative with limited linguistic performance bias

Many open questions:

- How does this transfer to speakers of non Indo-European languages?
- How about different interfaces (transcribing from paper)?
- What is the impact of font, colour, spacing and case?
- Could different syllabic patterns offer viable alternatives?
- Is removing some rare letters (like x) worth the entropy loss?
- What is the effect of chunking when typing spaces is not an issue?
- What makes codes memorable?

Thank you for your attention