# The Blizzard Challenge 2015

[1]*Kishore Prahallad,* [1]*Anandaswarup Vadapalli,* [1]*Sai Krishna Rallabandi,* [1]*Santosh Kesiraju,*
[2]*Hema Murthy,* [3]*T. Nagarajan,* [4]*Bira Singh,* [5]*Sajani T,* [6]*K Sreenivasa Rao,* [1]*Suryakanth V Gangashetty,*
[7]*Simon King,* [8]*Keiichi Tokuda,* [9]*Alan W Black*

[1] Speech and Vision Lab, IIIT Hyderabad, India
[2] Department of CSE, IIT Madras, India
[3] Department of IT, SSN College of Engineering, India
[4] CDAC Mumbai, India
[5] CDAC Trivandrum, India
[6] School of Information Technology, IIT Kharagpur, India
[7] Center for Speech Technology Research, University of Edinburgh, UK
[8] Department of Computer Science, Nagoya Institute of Technology, Japan
[9] Language Technologies Institute, Carnegie Mellon University, USA

## Abstract

The Blizzard challenge 2015 was the eleventh annual Blizzard challenge organised by the following group of institutions: IIIT Hyderabad, IIT Madras, SSN College of Engineering, CDAC Mumbai, CDAC Trivandrum and CDAC Kolkata with support and collaboration from DeitY, Government of India. This paper describes the tasks in Blizzard challenge 2015. The tasks consisted of data from six Indian languages : Bengali, Hindi, Malayalam, Marathi, Tamil and Telugu. Eight participants from around the world used the speech data provided as well as the corresponding text transcriptions in UTF-8, to build synthetic voices, which were then evaluated by means of listening tests.

**Index Terms**: Blizzard challenge, Speech synthesis, Evaluation of synthetic speech

## 1. Introduction

The Blizzard challenge, originally started by Profs. Black and Tokuda [1], is a well established challenge in the field of speech synthesis. [1–12] are summary papers describing the previous challenges. These resources can be found on the Blizzard challenge website [1]. This paper is a summary paper describing the Blizzard challenge 2015.

## 2. Nature of scripts and sounds of Indian languages

As a majority of Indian Languages use Indic scripts derived from the ancient Brahmi script, they share several orthographic patterns. The basic units of these scripts are called Aksharas, having the following properties: (i) An Akshara is an orthographic representation of one or more speech sounds in the concerned Indian language; (ii) Aksharas are mostly syllabic in nature; (iii) The canonical shapes of an Akshara are $V$, $CV$, $CCV$ and $CCCV$, and thus have a generalised form of $C^*V$, where $C$ stands for a consonant and $V$ stands for vowel.

Apart for sharing several orthographic patterns, most Indian languages (except a few such as English and Urdu) share a common phonetic base, i.e., they share a common set of speech sounds, in addition to a few more sounds individually. This common phonetic base consists of about 50 phones, including 15 vowels and 35 consonants. While all languages share a common phonetic base, some of the languages like Hindi and Marathi also share a common script called Devanagiri. Languages like Bengali, Malayalam, Tamil and Telugu have their own scripts.

The separation of these languages at the speech level can be attributed to the phonotactics of each language, rather than the scripts and speech sounds. Phonotactics are permissible combinations of phones that can co-occur in a language. This implies that the distribution of syllables in each language is different. Prosody (intonation, duration and prominence) associated with a syllable is another property that separates these languages significantly.

Another issue in the handling of Indian languages is that of digital representation of the scripts. With the advent of unicode, each letter of each language's script has it's own unique code point. This has standardised the representation of Aksharas and their rendering on the computer screen. However, the keying-in mechanism of these Aksharas has yet to be standardised. Due to this non-standardisation, the keying-in mechanism of Indian languages has to be addressed explicitly during the development of text processing modules in text-to-speech systems and user interfaces. In Blizzard challenge 2015 all Indian language text is in Unicode (UTF-8). To key-in Unicode, we link to Google transliterate[2], and use the transliteration scheme provided by them.

For further details on the nature of scripts and sounds of Indian languages please refer to [11].

## 3. Blizzard challenge 2015 tasks

### 3.1. Data used

Speech and text data for six Indian languages: i) Bengali, ii) Hindi, iii) Malayalam, iv) Marathi, v) Tamil and vi) Telugu were released. The speech data for Hindi, Tamil and Telugu was

[1] http://www.festvox.org/blizzard/

[2] http://www.google.com/inputtools/try/

4 hours each, while for the remaining three languages it was 2 hours each. For all six languages the speech data was sampled at 16 kHz and recorded by professional speakers in a high quality studio environment. Along with the speech data the corresponding text was provided in UTF-8 format. No other information, like segment labels was provided as part of the challenge. However, there was no restriction on the participants to learn/use information like phonesets or labels from other resources.

### 3.2. Tasks

The Blizzard challenge 2015 consisted of two tasks, a hub task and a spoke task.

- Hub task 2015-IH1: Participants were asked to build one voice in each language from the provided data, in accordance of the rules of the challenge. The subtasks were numbered from IH1.1 to IH1.6 corresponding to the six languages: IH1.1 (Bengali), IH1.2 (Hindi), IH1.3 (Malayalam), IH1.4 (Marathi), IH1.5 (Tamil) and IH1.6 (Telugu).
- Spoke task 2015-IH2: Participants had to synthesize multilingual sentences containing Indian language text as well as English. The subtasks were numbered from IH2.1 to IH2.6 corresponding to the six languages: IH2.1 (Bengali), IH2.2 (Hindi), IH2.3 (Malayalam), IH2.4 (Marathi), IH2.5 (Tamil) and IH2.6 (Telugu).

For the IH1 task (hub task), the synthetic voices were evaluated through listening tests on the following test data (for each Indian language)

- Read speech (RD) - 50 distinct sentences, not a part of the training data
- Semantically unpredictable sentences (SUS) - 50 distinct sentences not a part of RD/training data

The SUS sentences were prepared in the following manner. 50 sentences in each language were randomly selected, and POS tagging was performed on these sentences. The words in each sentence were then reordered as Subject Object Verb Conjuction Subject Object Verb to generate the SUS sentence.

For the IH2 task (spoke task), the systems were evaluated through listening tests by synthesizing the following test data (for each Indian language + English combination)

- Multilingual sentences (ML) - 50 distinct sentences containing both Indian language as well as English words.

No language tags were provided in the ML sentences. The participants were expected to identify the language from the Unicode code point.

### 3.3. Participants in the challenge

The participants in the Blizzard challenge 2015 consisted of the eight participants listed in Table 1. To annonimyze the results, the systems are identified using letters, with A denoting natural speech, B denoting the baseline system and C to J denoting the systems submitted by the participants in the challenge. Each participant could submit as many systems as they wished.

### 3.4. Baseline systems

Baseline voices were built for each language using FestVox [13] in the unit selection framework. For this purpose, the FestVox scripts specific to building Indic voices[3] were used.

Table 1: Participants in Blizzard challenge 2015

| Short name | Details | Synthesis method |
|---|---|---|
| NATURAL | Natural speech | |
| BASELINE | Baseline system | Concatenative |
| CMU | Carnegie Mellon University,<br>USA | SPS - Random Forests |
| DONLAB | Indian Institute of Technology - Madras,<br>India | SPS-HTS |
| IIITH | International Institute of Information<br>Technology - Hyderabad, India | Concatenative |
| NELSLIP | National Engineering Laboratory for<br>Speech and Language Information Processing,<br>University of Science and Technology, China | Hybrid |
| NITECH | Nagoya Institute of Technology,<br>Japan | SPS-HTS |
| CSTR_HELSINKI | Center for Speech Technology Research,<br>Univ. of Endinburgh, U.K<br>and Univ. of Helsinki, Finland | SPS-DNN |
| IRISA | Institute de Recherche en Informatique et<br>Systemes Aleatoires, France | Concatenative |
| NAIST | Nara Institute of Science and Technology,<br>Japan | SPS-HTS |

## 4. Evaluation

The participants were asked to synthesize the complete test set, out of which a subset was used in the listening tests. The listening tests for IH1.1 - IH1.6 consisted of ten sections while the listening tests for IH2.1 - IH2.6 consisted of five sections. The different sections of the listening tests are described below.

- Listening tests for IH1.1 - IH1.6
  1. two sections for similarity (one section using RD and one section using SUS)
  2. seven sections for naturalness (four sections using RD and three sections using SUS)
  3. one section for intelligibility using SUS
- Listening tests for IH2.1 - IH2.6
  1. one section for similarity
  2. four sections for naturalness

The methodology of scoring in the various sections of the listening tests are described below.

- **Similarity** : The listener plays a few samples of the original speaker and one synthetic sample. The listener then chooses a response that represented how similar the synthetic voice sounded as compared to the original speakers voice on a scale from

  1 : Sounds like a totally different person
  to
  5 : Sounds exactly like the same person

- **Naturalness** : The listener listenes to a sample of synthetic speech and chooses a score which represents how natural or unnatural the sentence sounded on a scale of

  1 : Completely Unnatural
  to
  5 : Completely Natural

- **Intelligibility** : Listeners listen to an utterance and type in what they hear. Word Error Rate (WER) is computed in the same manner as it is computed for speech recognition tasks.

For the list of changes made in the evaluation portal to enable the conduct of listening tests in Indian languages, please refer to [11]

[3] http://festvox.org/bsv/x3528.html

# 5. Results

All the listening tests conducted for Blizzard challenge 2015 tasks, were conducted using paid listeners only. Table 2 shows the statistics of the listeners for the different tasks.

Table 2: User statistics for the Blizzard 2015 tasks

| Task | Paid Users |
|---|---|
| IH1.1 + IH2.1 | 48 |
| IH1.2 + IH2.2 | 69 |
| IH1.3 + IH2.3 | 72 |
| IH1.4 + IH2.4 | 69 |
| IH1.5 + IH2.5 | 70 |
| IH1.6 + IH2.6 | 70 |

### 5.1. Results

Tables 3 to 8 show the mean MOS of naturalness and similarity on RD, SUS and ML for all six languages (Bengali, Hindi, Malayalam, Marathi, Tamil and Telugu). In all the Tables, the maximum value in each column in represented in **bold** and the mininum value in each column is represented in *italics*.

For the six languages in the IH1 hub task (IH1.1 - IH1.6), Figures 1 to 6 and Figures 7 to 12 show the similarity and naturalness results on RD and SUS respectively.

The intelligibility results for the hub task (IH1.1 - IH1.6) are shown inFigures 13 to 18.

For the spoke task (IH2.1 - IH2.6), Figures 19 to 24 show the similarity and naturalness results on ML.

For a detailed discussion of the results, please refer to the papers describing each system submitted by individual participants, available on the Blizzard Challenge website.

# 6. Conclusions

The conclusions drawn from the results of the Blizzard challenge 2015 are:

1. High quality audio recordings provided decent performances by all systems
2. There does not seem to be much utility in computing WER as a measure of intelligibility for Indian languages, as seen from the globally bad WER numbers across all systems and all languages. Such high WER scores can be explained by the following:
    - Native speakers are not used to typing Indian language scripts as there is no standard keyboard layout.
    - The Google transliteration APIs that are used to key-in Indian language text during evaluation, require a space to be pressed before the ASCII character is changed to UTF8 script. The space is often missed by testers, resulting in errors.
    - WER computation is done using a binary match, which gives a lot of errors due to spelling mistakes while keying-in. This is especially noticeable in the case of long vowels vs. short vowels.
3. There is a requirement for different measures for speaker similarity and mean-opinion scores especially for Indian languages.

# 7. Acknowledgements

Thanks to all the people who participated in the evaluations. Thanks also to the previous organizers of the Blizzard challege for providing the scripts for conducting the listening tests.

# 8. References

[1] A. W. Black and K. Tokuda, “The Blizzard Challenge - 2005 : Evaluating corpus-based speech synthesis on common datasets,” in *Proceedings of Intespeech 2005*, Lisbon, 2005.

[2] C. L. Bennett, “Large scale evaluation of corpus-based synthesizers : Results and lessons from the Blizzard Challenge 2005,” in *Proceedings of Interspeech 2005*, 2005.

[3] C. L. Bennett and A. W. Black, “The Blizzard Challenge 2006,” in *Blizzard Challenge Workshop, Interspeech 2006 - ICSLP satellite event*, 2006.

[4] M. Frazer and S. King, “The Blizzard Challenge 2007,” in *Proceedings Blizzard Workshop 2007 (in Proceedings SSW6)*, 2007.

[5] V. Karaiskos, S. King, R. Clark, and C. Mayo, “The Blizzard Challenge 2008,” in *Proceedings Blizzard Workshop 2008*, 2008.

[6] S. King and V. Karaiskos, “The Blizzard Challenge 2009,” in *Proceedings Blizzard Workshop 2009*, 2009.

[7] ——, “The Blizzard Challenge 2010,” in *Proceedings Blizzard Workshop 2010*, 2010.

[8] ——, “The Blizzard Challenge 2011,” in *Proceedings Blizzard Workshop 2011*, 2011.

[9] ——, “The Blizzard Challenge 2012,” in *Proceedings Blizzard Workshop 2012*, 2012.

[10] ——, “The Blizzard Challenge 2013,” in *Proceedings Blizzard Workshop 2013*, 2013.

[11] K. Prahallad, A. Vadapalli, N. Elluru, G. Mantena, B. Pulugundla, P. Bhaskararao, H. A. Murthy, S. King, V. Karaiskos, and A. W. Black, “The Blizzard Challenge 2013 – Indian Language Tasks,” in *Proceedings Blizzard Workshop 2013*, 2013.

[12] K. Prahallad, A. Vadapalli, S. Kesiraju, H. A. Murthy, S. Lata, T. Nagarajan, M. Prasanna, H. Patil, A. K. Sao, S. King, A. W. Black, and K. Tokuda, “The Blizzard Challenge 2014,” in *Proceedings Blizzard Workshop 2014*, 2014.

[13] A. W. Black and K. Lenzo, “Building voices in the festival speech synthesis system,” 2002, available Online: http://festvox.org/bsv.

[14] R. Clark, M. Podsiadlo, M. Fraser, C. Mayo, and S. King, “Statistical analysis of the Blizzard Challenge 2007 listening test results,” in *Proceeding Blizzard Workshop 2007 (in Proceedings SSW6)*, 2007.

Table 3: Mean MOS scores for IH1.1 (Bengali)

| System ID | RD (Mean MOS) | | SUS (Mean MOS) | | ML (Mean MOS) | |
|---|---|---|---|---|---|---|
| | NAT | SIM | NAT | SIM | NAT | SIM |
| A | 4.7 | 4.5 | 4.6 | 4.7 | 4.7 | 4.6 |
| B | 2.2 | 2.5 | 2.7 | **3.7** | 1.8 | *1.8* |
| C | 2.9 | 2.2 | **3.1** | *2.2* | 2.6 | 2.1 |
| D | 3.0 | 2.2 | 3.0 | 3.0 | 2.6 | 2.8 |
| E | **3.4** | 3.7 | 2.7 | 3.3 | **3.8** | **4.1** |
| F | *2.0* | *1.7* | *1.8* | 2.6 | *1.6* | 2.1 |
| G | 2.5 | **3.1** | 2.8 | 3.1 | 2.2 | 2.2 |
| H | 2.6 | 2.4 | 2.6 | 2.5 | - | - |
| I | 2.7 | 2.9 | 2.1 | 2.3 | - | - |
| J | 2.6 | 2.7 | 2.5 | 2.8 | - | - |

Table 4: Mean MOS scores for IH1.2 (Hindi)

| System ID | RD (Mean MOS) | | SUS (Mean MOS) | | ML (Mean MOS) | |
|---|---|---|---|---|---|---|
| | NAT | SIM | NAT | SIM | NAT | SIM |
| A | 4.7 | 4.5 | 4.4 | 4.4 | 4.7 | 4.5 |
| B | 3.2 | 2.6 | 2.6 | 3.5 | *1.8* | *1.8* |
| C | 3.5 | *2.4* | 3.3 | *2.0* | **3.2** | 2.2 |
| D | 2.8 | 2.7 | 2.7 | 2.2 | 2.3 | 2.0 |
| E | 2.6 | 2.9 | 3.0 | 2.8 | **3.2** | **3.5** |
| F | **3.9** | **4.3** | **3.9** | **3.9** | 2.9 | 3.1 |
| G | *2.3* | 2.8 | *2.4* | 2.9 | 2.0 | 2.2 |
| H | 2.8 | 2.7 | 3.0 | 2.6 | - | - |
| I | 2.8 | 3.5 | 3.2 | 3.3 | - | - |
| J | 3.3 | 3.3 | 3.1 | 2.2 | - | - |

Table 5: Mean MOS scores for IH1.3 (Malayalam)

| System ID | RD (Mean MOS) | | SUS (Mean MOS) | | ML (Mean MOS) | |
|---|---|---|---|---|---|---|
| | NAT | SIM | NAT | SIM | NAT | SIM |
| A | 4.3 | 4.6 | 4.3 | 4.2 | 4.4 | 4.2 |
| B | *1.6* | *1.8* | *1.9* | *2.1* | *1.9* | 2.6 |
| C | 2.6 | 2.3 | 2.8 | *2.1* | 2.4 | 2.2 |
| D | 2.3 | 2.1 | 2.4 | 2.2 | 2.2 | 2.1 |
| E | 2.3 | 2.9 | 2.7 | 2.6 | **3.6** | **3.2** |
| F | **2.9** | 2.3 | 2.5 | 2.3 | 2.7 | 2.6 |
| G | 1.7 | 2.3 | 2.0 | *2.1* | *1.9* | *2.0* |
| H | 2.1 | 2.0 | 2.1 | 2.2 | - | - |
| I | 2.7 | **3.0** | **2.9** | **3.2** | - | - |
| J | **2.9** | 2.0 | 2.3 | 2.9 | - | - |

Table 6: Mean MOS scores for IH1.4 (Marathi)

| System ID | RD (Mean MOS) | | SUS (Mean MOS) | | ML (Mean MOS) | |
|---|---|---|---|---|---|---|
| | NAT | SIM | NAT | SIM | NAT | SIM |
| A | 4.6 | 4.4 | 4.5 | 4.3 | 4.8 | 4.3 |
| B | 2.7 | 2.3 | 2.5 | 2.7 | 2.2 | 2.2 |
| C | 2.5 | 2.3 | 2.7 | *1.9* | 2.6 | 1.6 |
| D | 3.0 | **3.0** | 2.9 | **3.4** | 2.6 | **3.1** |
| E | 3.0 | 2.9 | **3.3** | **3.4** | **3.4** | 2.7 |
| F | 3.2 | **3.0** | 3.2 | 2.8 | 2.9 | 2.7 |
| G | *2.2* | 2.5 | *2.2* | 2.2 | *2.1* | *2.1* |
| H | 2.9 | *2.1* | 2.7 | 2.4 | - | - |
| I | - | - | - | - | - | - |
| J | **3.3** | 2.8 | 2.9 | 2.4 | - | - |

Table 7: Mean MOS scores for IH1.5 (Tamil)

| System ID | RD (Mean MOS) | | SUS (Mean MOS) | | ML (Mean MOS) | |
|---|---|---|---|---|---|---|
| | NAT | SIM | NAT | SIM | NAT | SIM |
| A | 4.7 | 4.6 | 4.6 | 4.6 | 4.7 | 4.2 |
| B | *2.2* | *1.8* | *2.2* | *1.9* | *2.2* | *2.2* |
| C | 2.8 | 2.2 | 3.3 | 2.8 | 2.9 | 2.8 |
| D | 2.6 | 1.9 | 2.6 | 2.0 | 2.3 | 1.7 |
| E | 2.5 | 2.7 | 3.0 | 2.6 | **4.0** | **3.3** |
| F | **3.6** | 2.7 | 3.2 | 2.5 | 3.3 | 2.6 |
| G | 2.4 | 2.3 | 2.3 | 3.1 | 2.6 | 2.4 |
| H | 3.0 | 2.6 | **3.7** | 2.7 | - | - |
| I | 3.2 | **3.6** | 3.0 | **3.4** | - | - |
| J | 2.7 | 2.6 | 3.0 | 2.3 | - | - |

Table 8: Mean MOS scores for IH1.6 (Telugu)

| System ID | RD (Mean MOS) | | SUS (Mean MOS) | | ML (Mean MOS) | |
|---|---|---|---|---|---|---|
| | NAT | SIM | NAT | SIM | NAT | SIM |
| A | 4.8 | 4.5 | 4.5 | 3.3 | 4.8 | 3.8 |
| B | *1.9* | 2.1 | *1.8* | 2.1 | *2.0* | 2.1 |
| C | 2.6 | *1.3* | 2.5 | *1.4* | 2.5 | 1.5 |
| D | 2.1 | 2.0 | 2.5 | 2.6 | 2.6 | 2.3 |
| E | 2.8 | 2.9 | 2.7 | 2.5 | **2.9** | **2.5** |
| F | **3.5** | 2.5 | 2.5 | 2.0 | 2.6 | 2.2 |
| G | 2.1 | 3.1 | 2.1 | 2.2 | 2.4 | *1.4* |
| H | 3.0 | 2.4 | **3.0** | **3.4** | - | - |
| I | 2.9 | **4.2** | 2.1 | 1.9 | - | - |
| J | **3.5** | 2.7 | 2.7 | 2.0 | - | - |

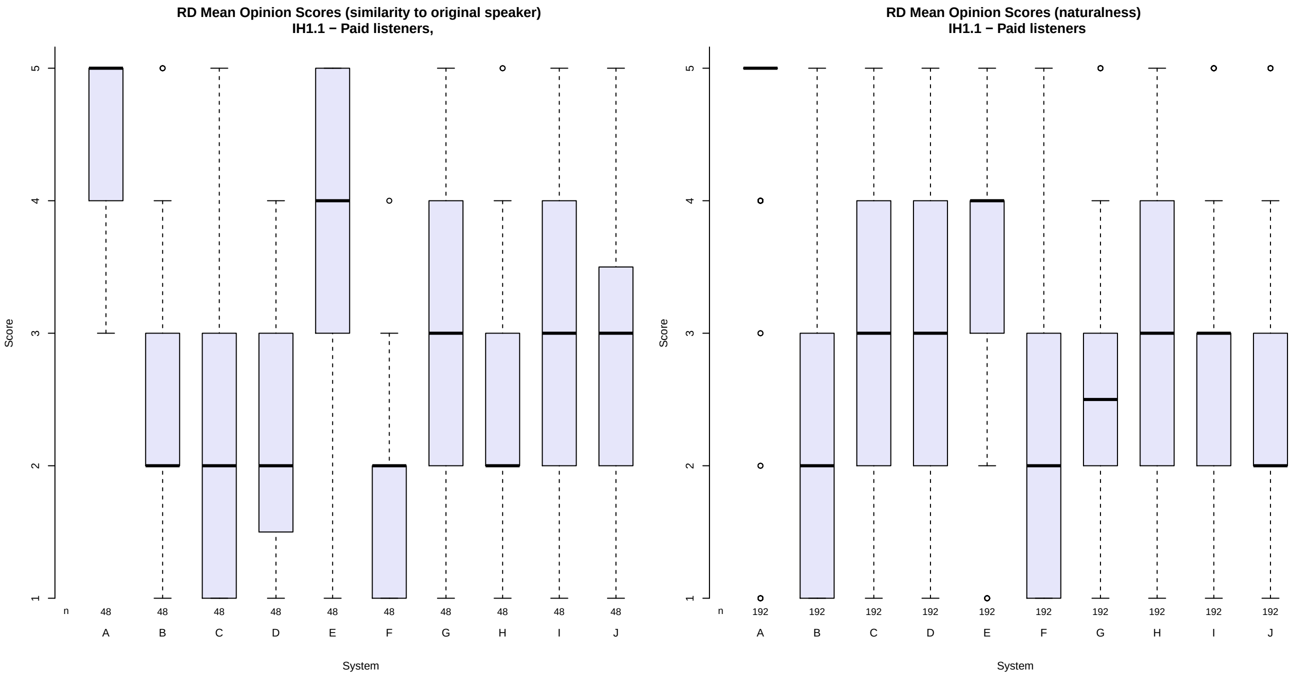

Figure 1: Similarity and Naturalness results on RD for IH1.1 (Bengali)

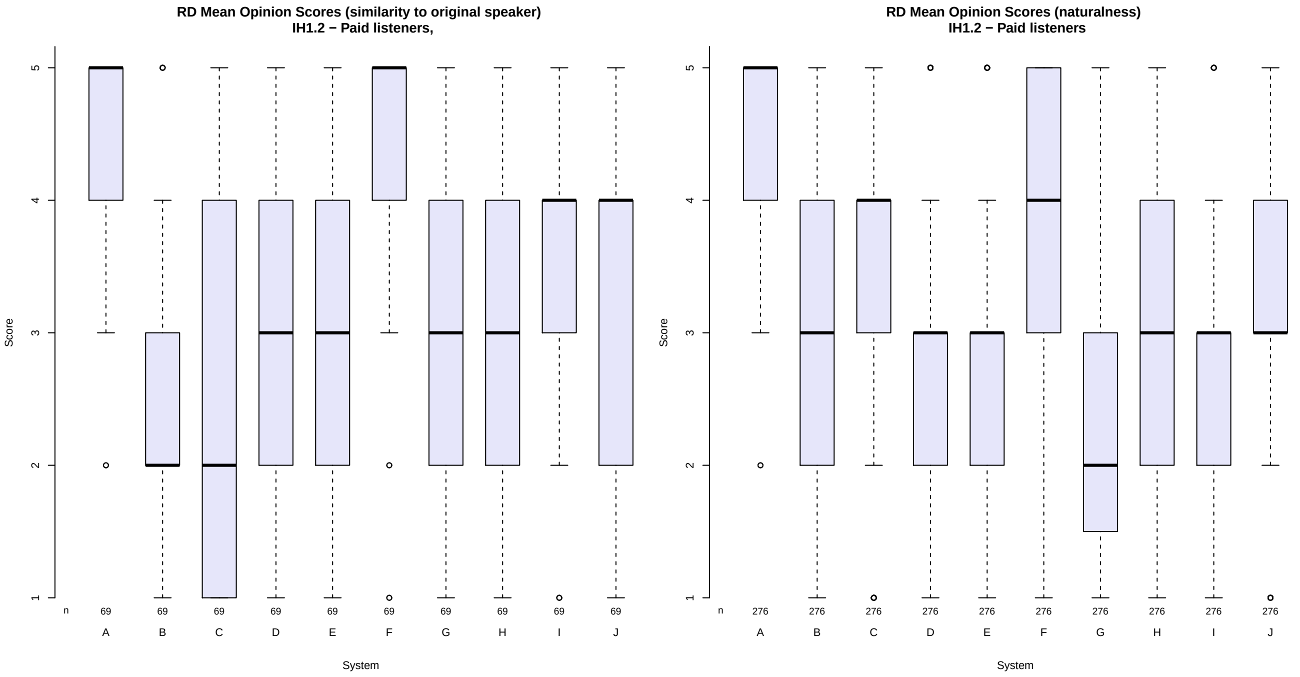

Figure 2: Similarity and Naturalness results on RD for IH1.2 (Hindi)

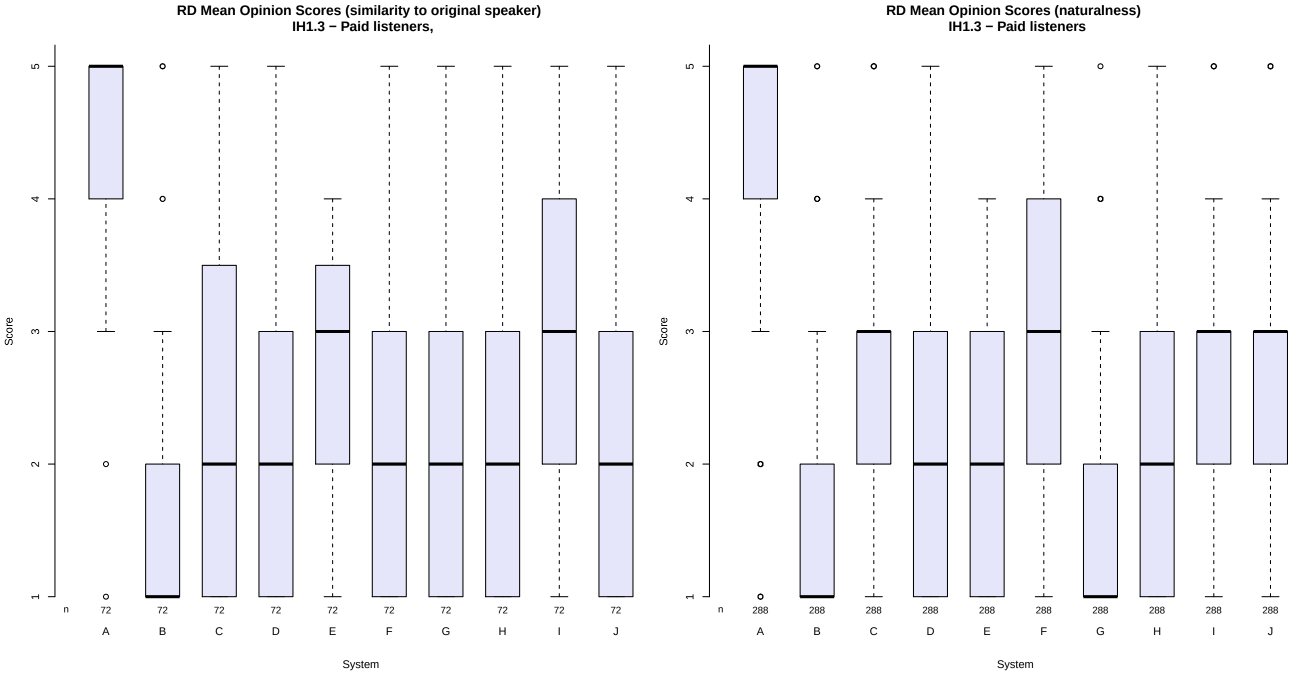

Figure 3: Similarity and Naturalness results on RD for IH1.3 (Malayalam)

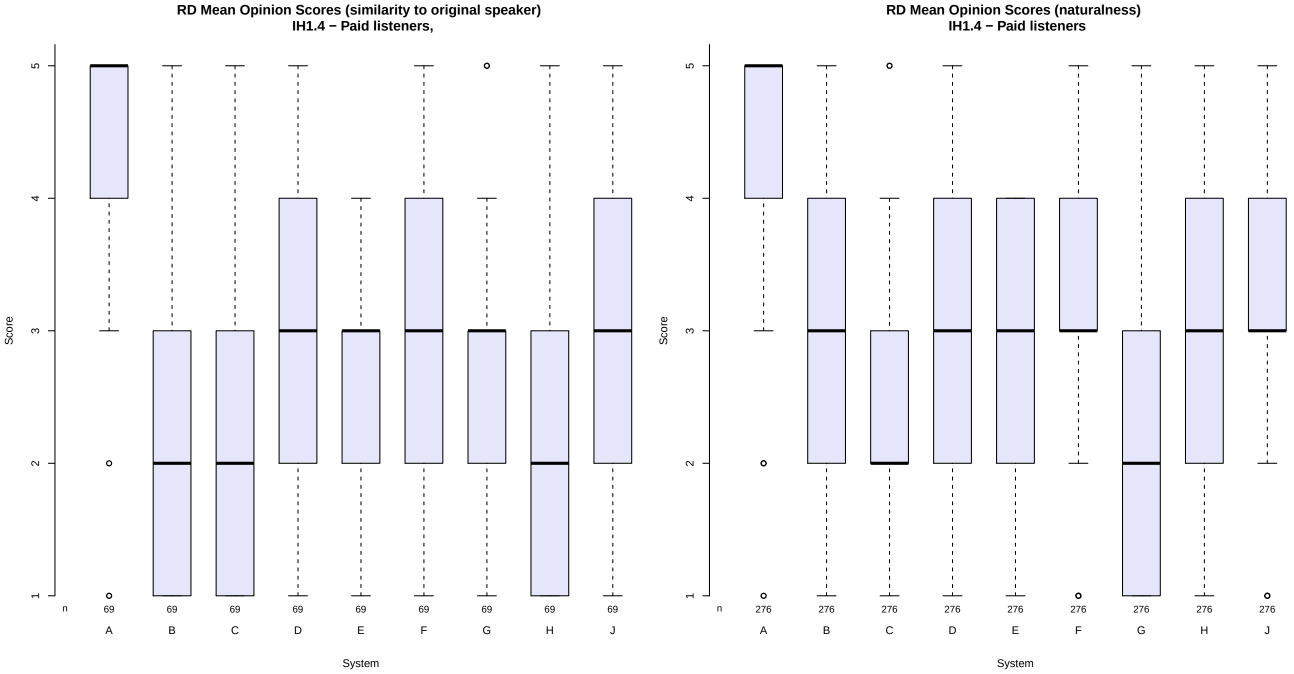

Figure 4: Similarity and Naturalness results on RD for IH1.4 (Marathi)

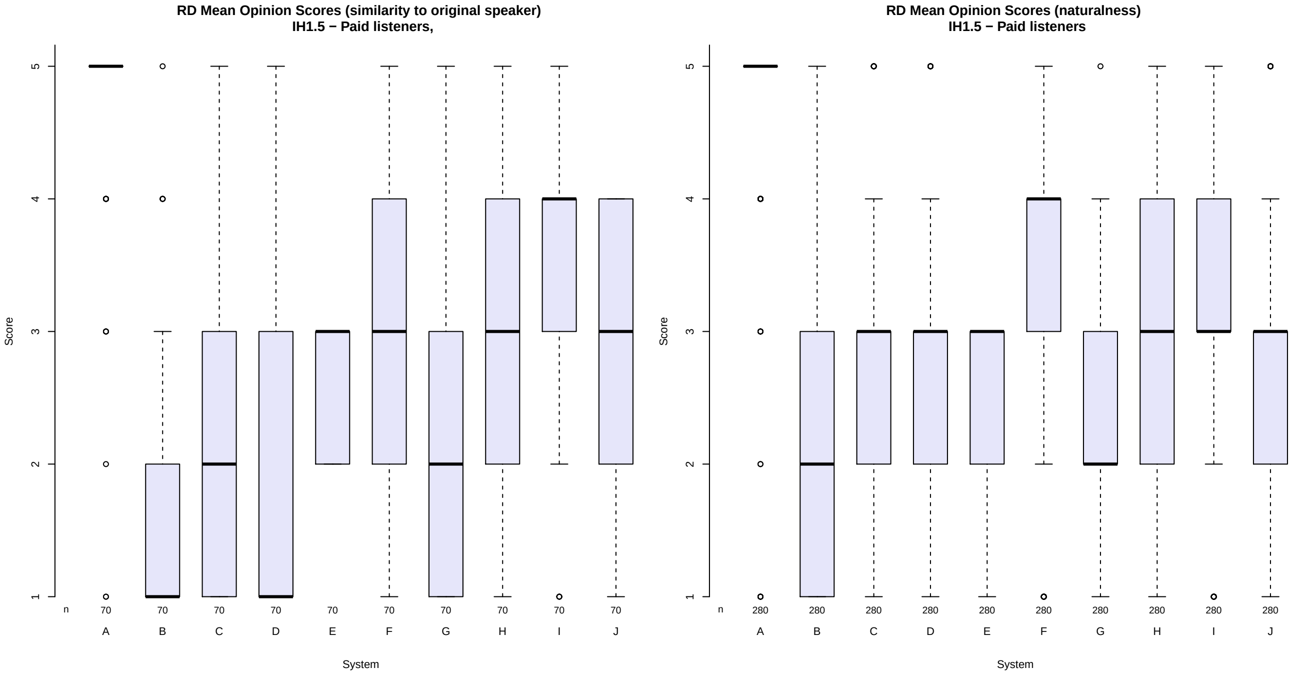

Figure 5: Similarity and Naturalness results on RD for IH1.5 (Tamil)

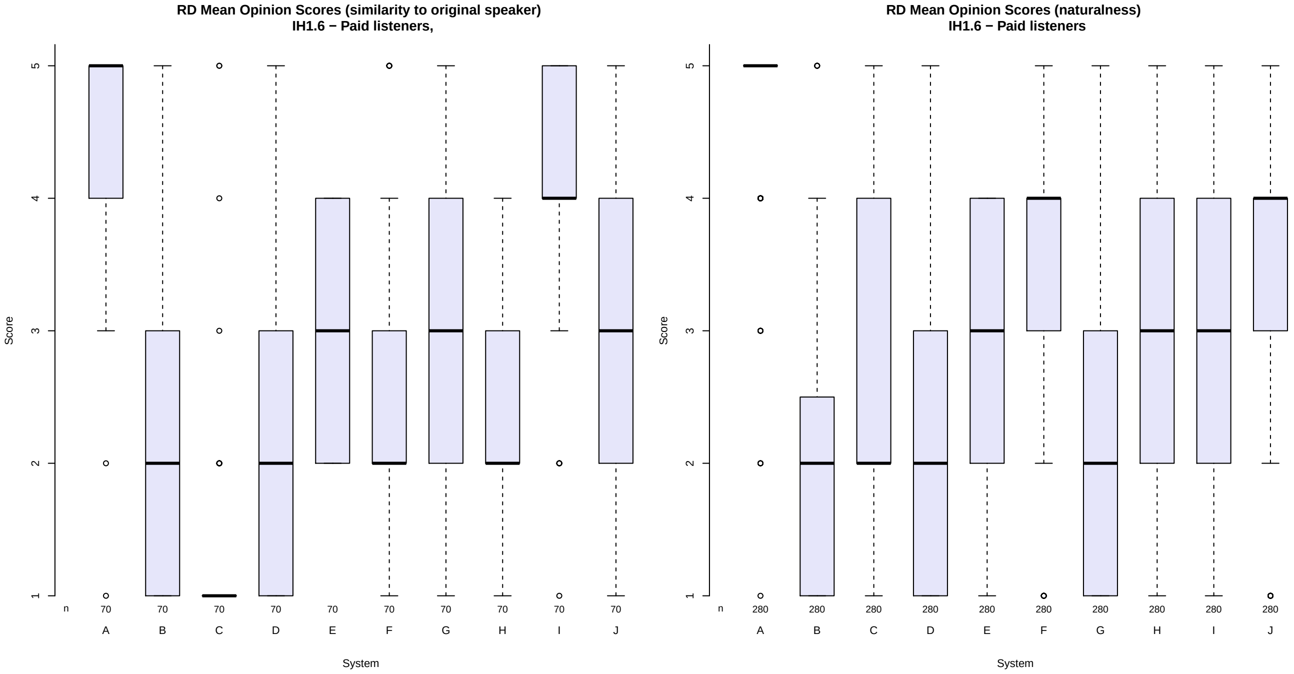

Figure 6: Similarity and Naturalness results on RD for IH1.6 (Telugu)

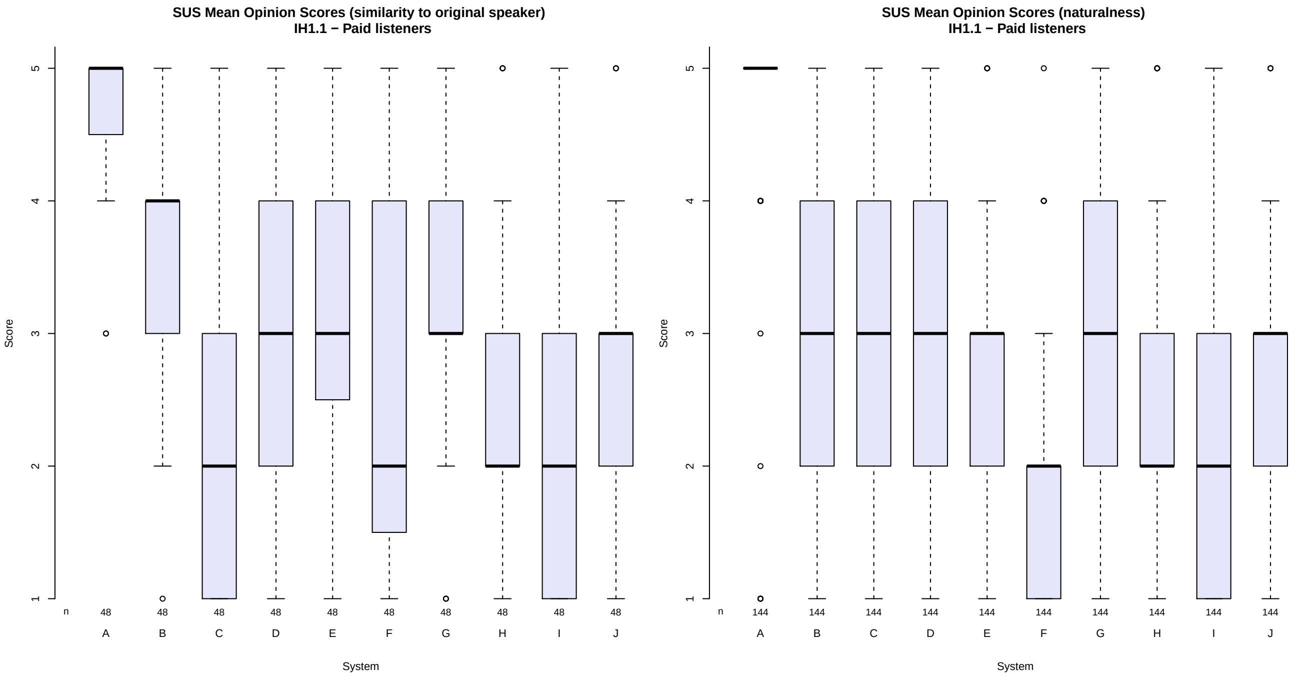

Figure 7: Similarity and Naturalness results on SUS for IH1.1 (Bengali)

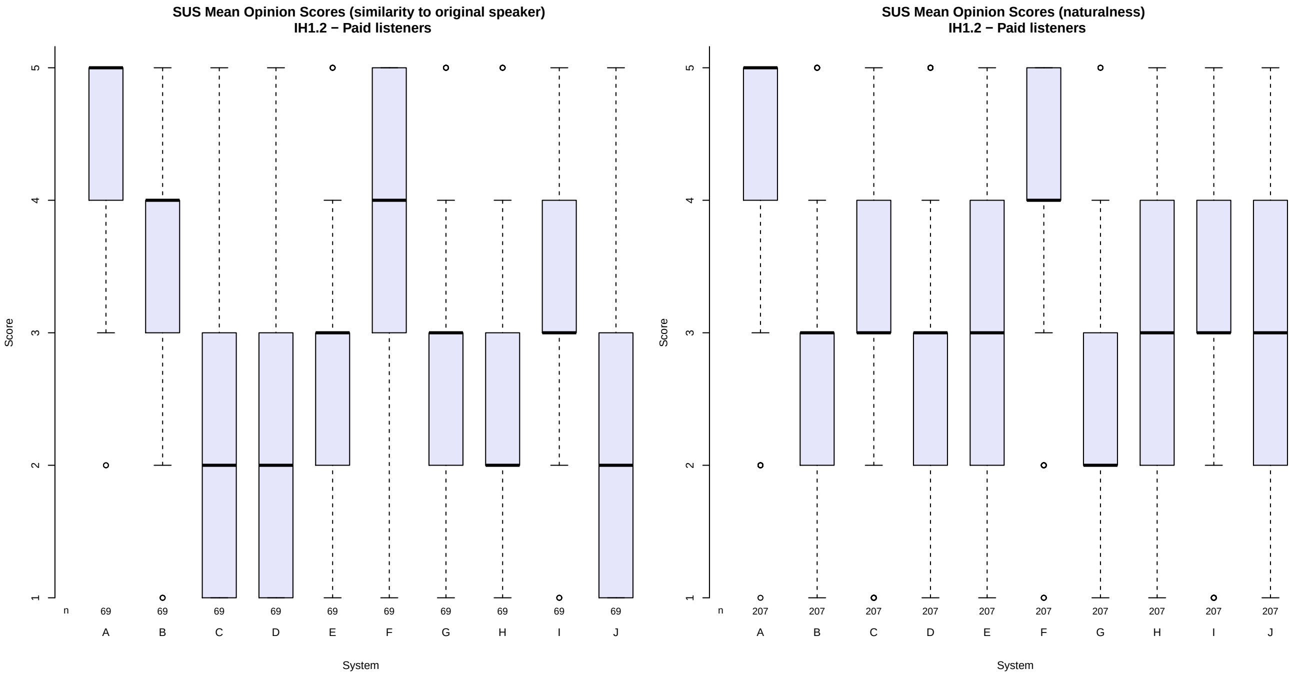

Figure 8: Similarity and Naturalness results on SUS for IH1.2 (Hindi)

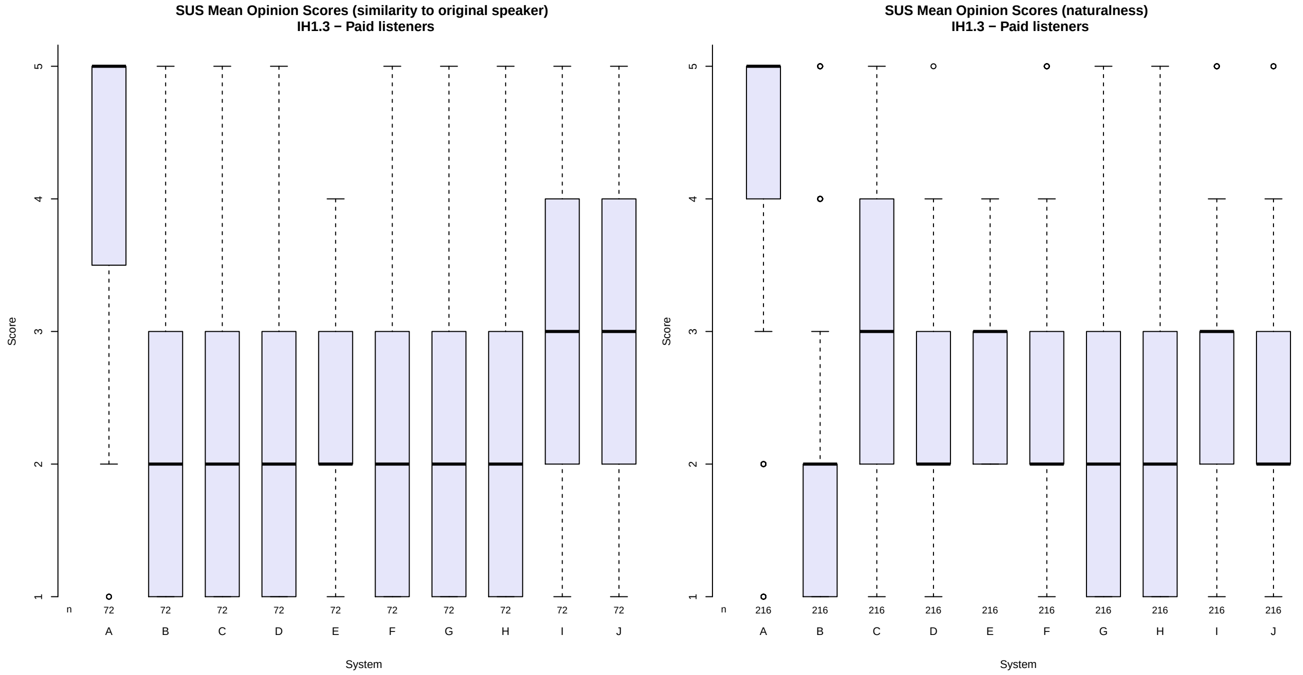

Figure 9: Similarity and Naturalness results on SUS for IH1.3 (Malayalam)

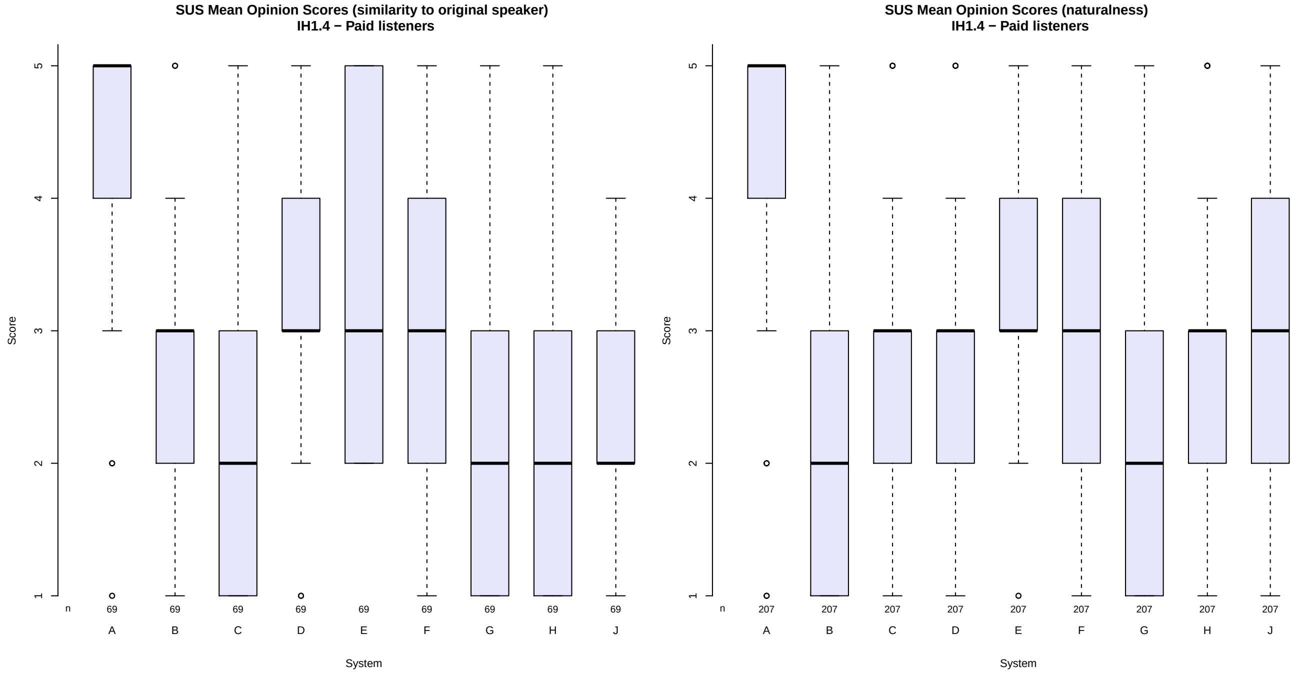

Figure 10: Similarity and Naturalness results on SUS for IH1.4 (Marathi)

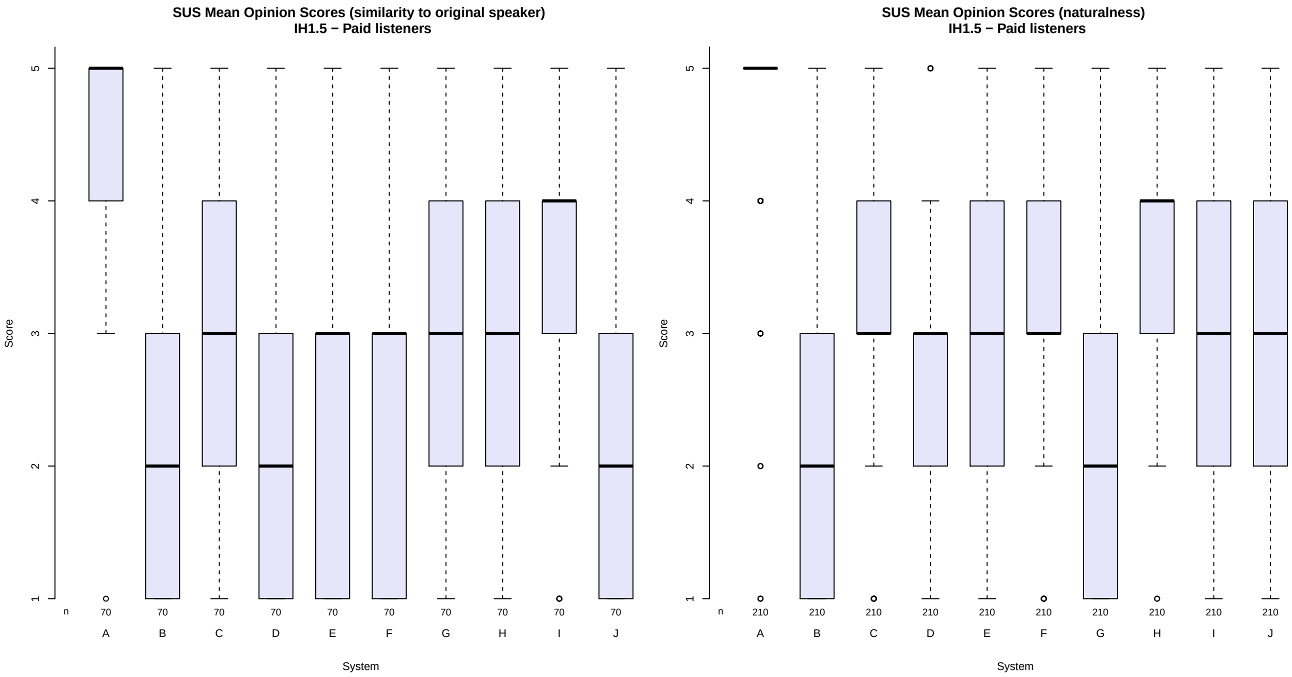

Figure 11: Similarity and Naturalness results on SUS for IH1.5 (Tamil)

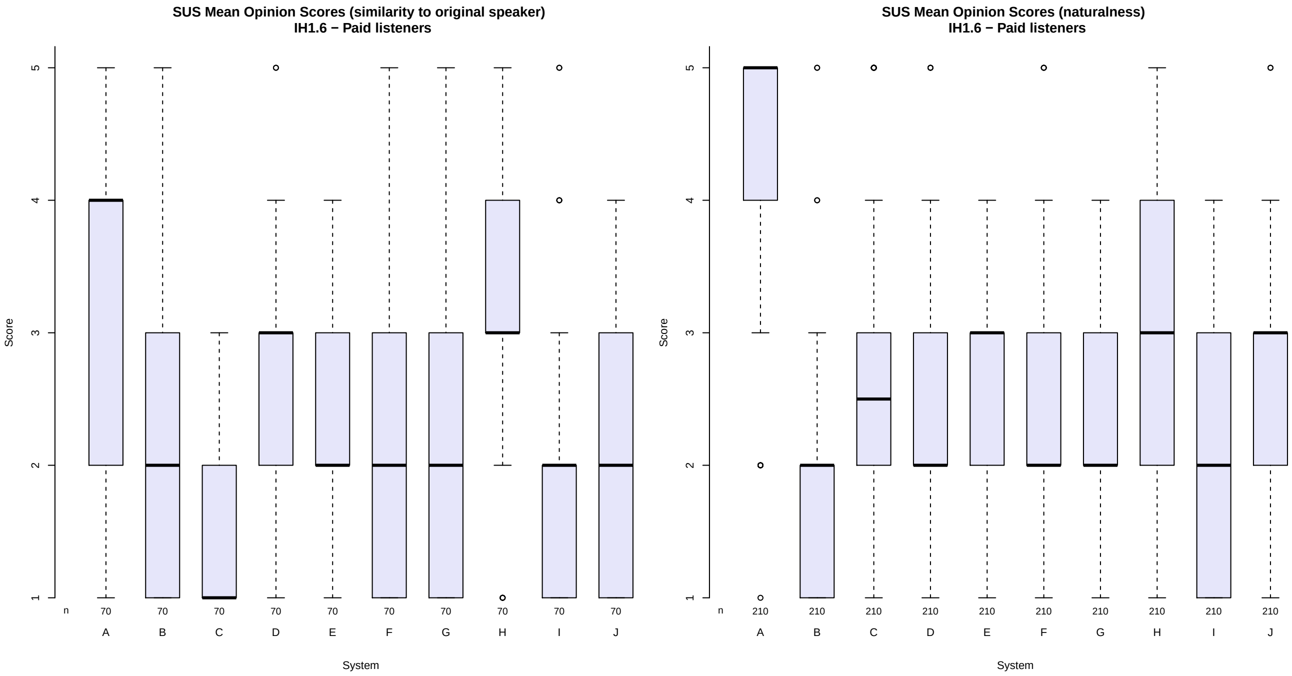

Figure 12: Similarity and Naturalness results on SUS for IH1.6 (Telugu)

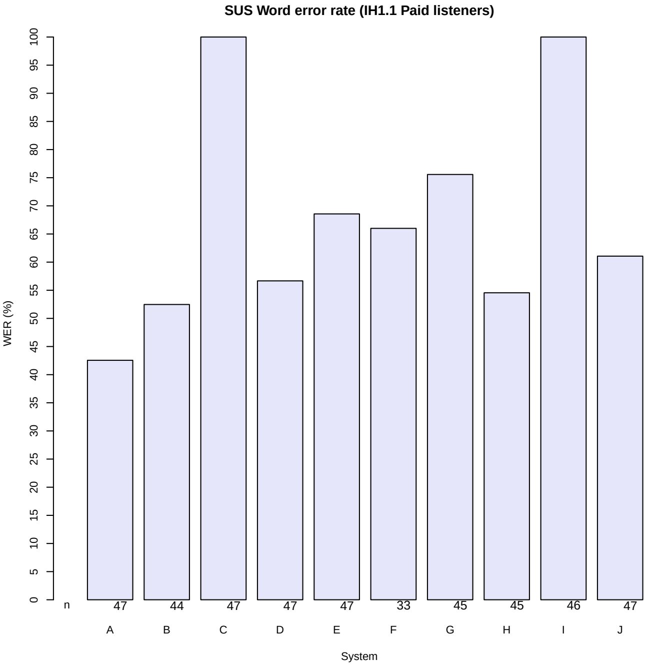

Figure 13: Intelligibility results on SUS for IH1.1 (Bengali)

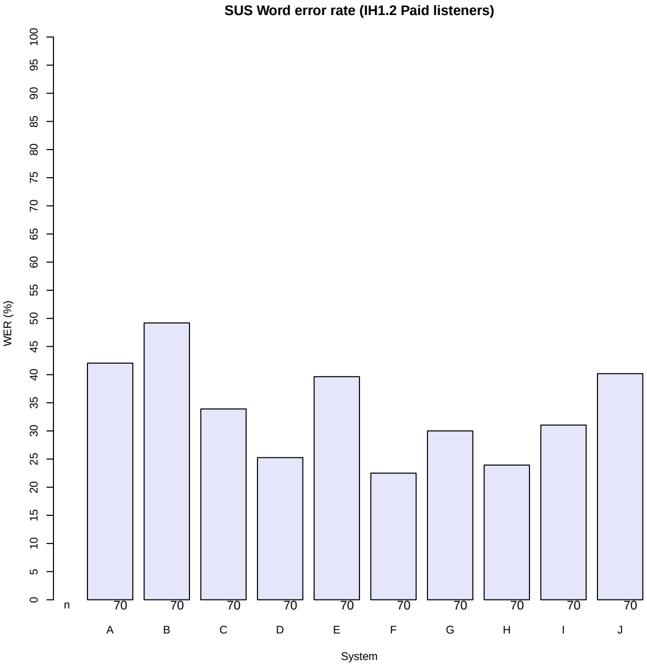

Figure 14: Intelligibility results on SUS for IH1.2 (Hindi)

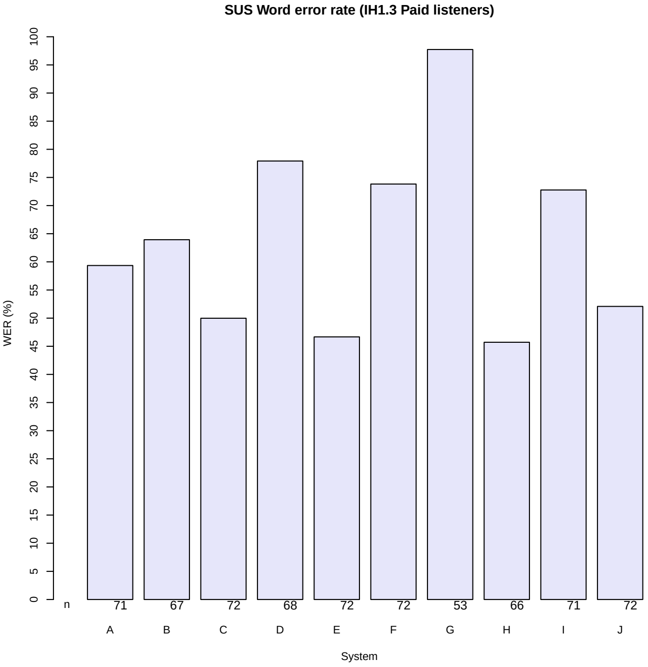

Figure 15: Intelligibility results on SUS for IH1.3 (Malayalam)

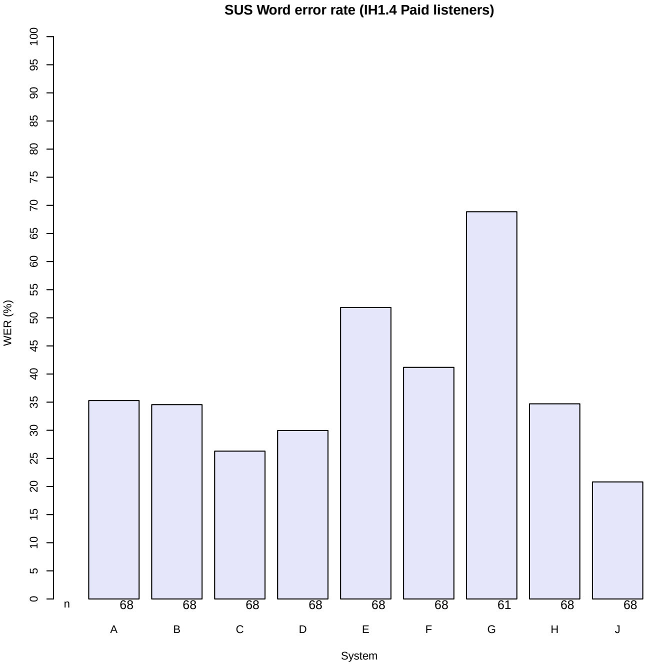

Figure 16: Intelligibility results on SUS for IH1.4 (Marathi)

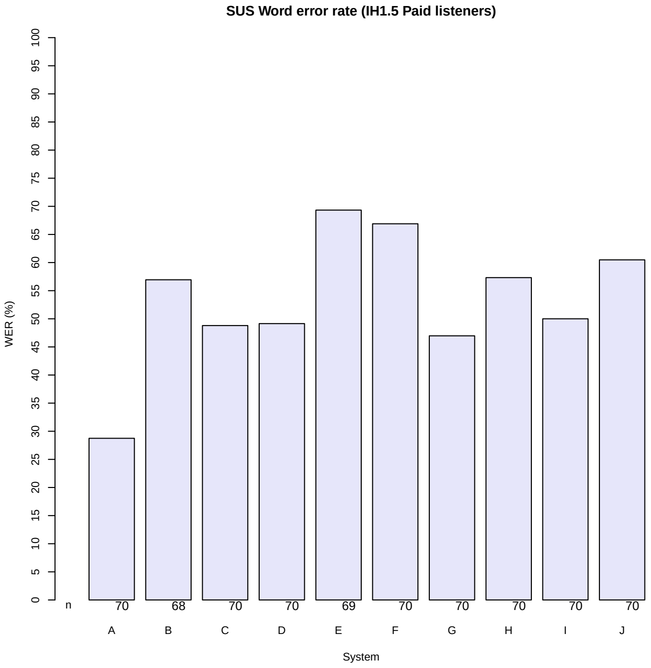

Figure 17: Intelligibility results on SUS for IH1.5 (Tamil)

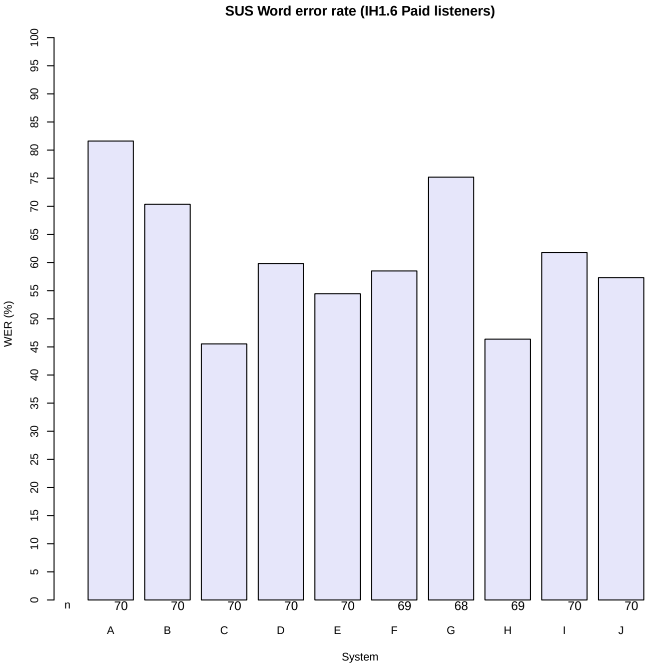

Figure 18: Intelligibility results on SUS for IH1.6 (Telugu)

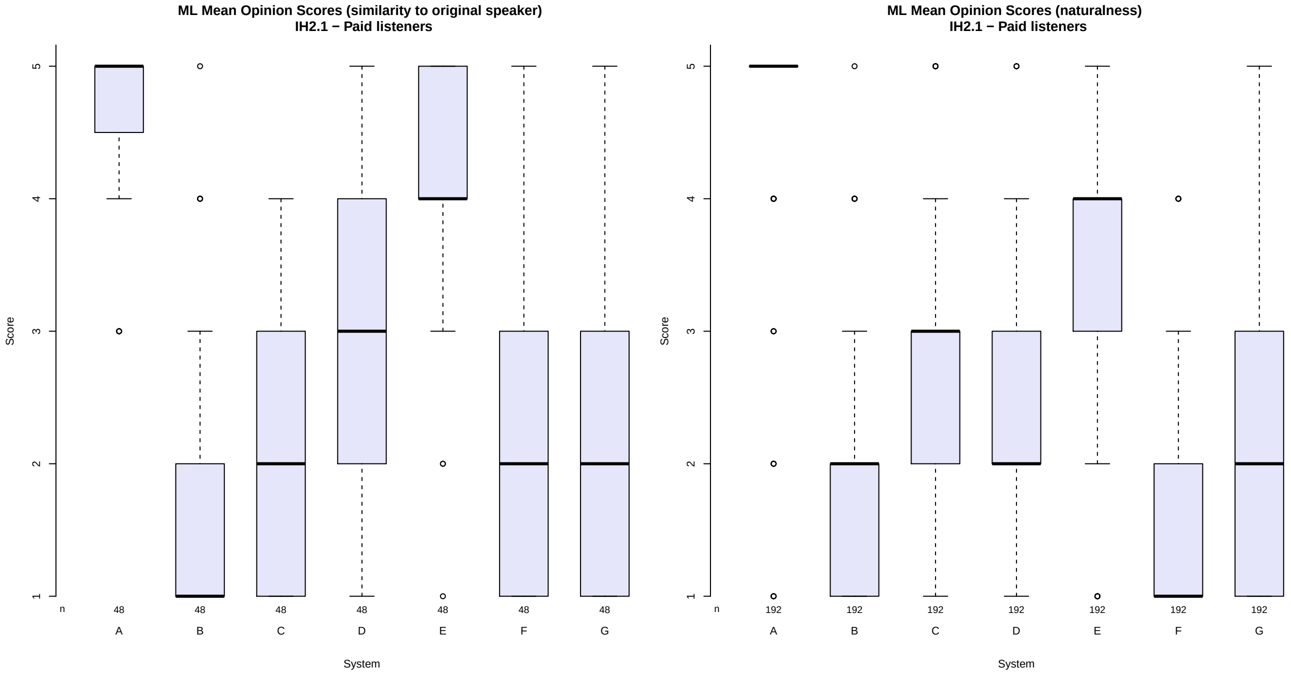

Figure 19: Similarity and Naturalness results on ML for IH2.1 (Bengali)

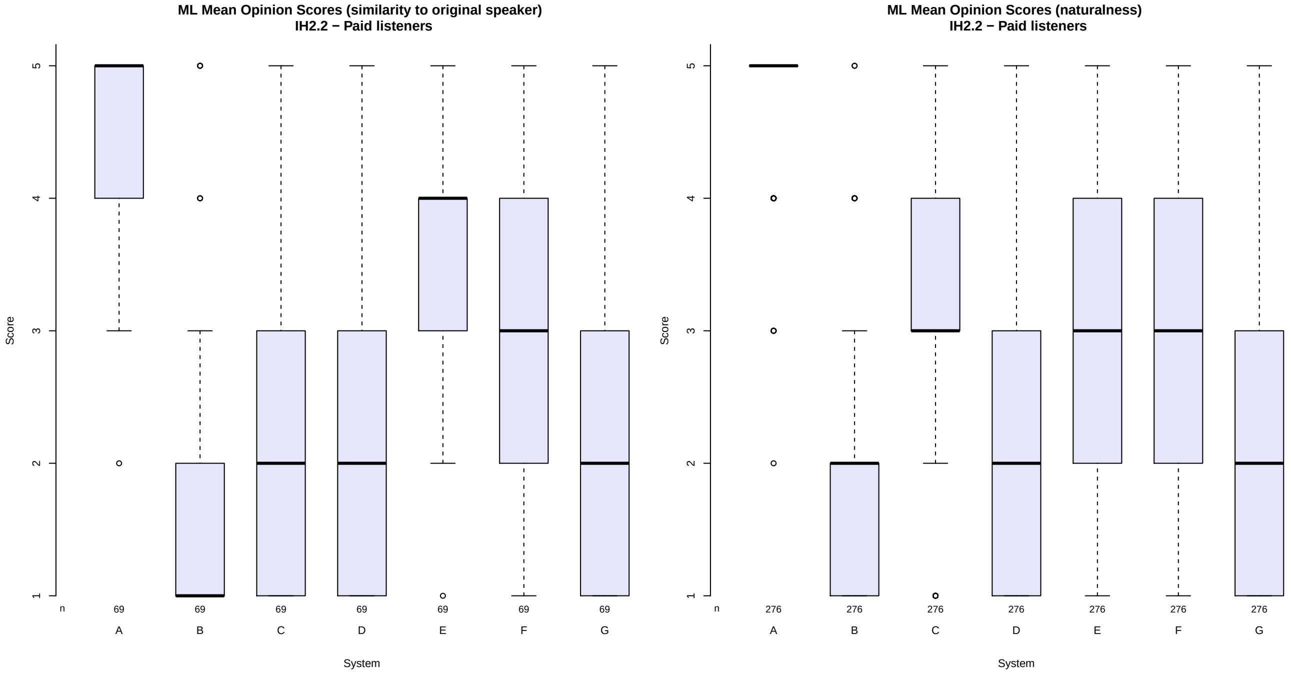

Figure 20: Similarity and Naturalness results on ML for IH2.2 (Hindi)

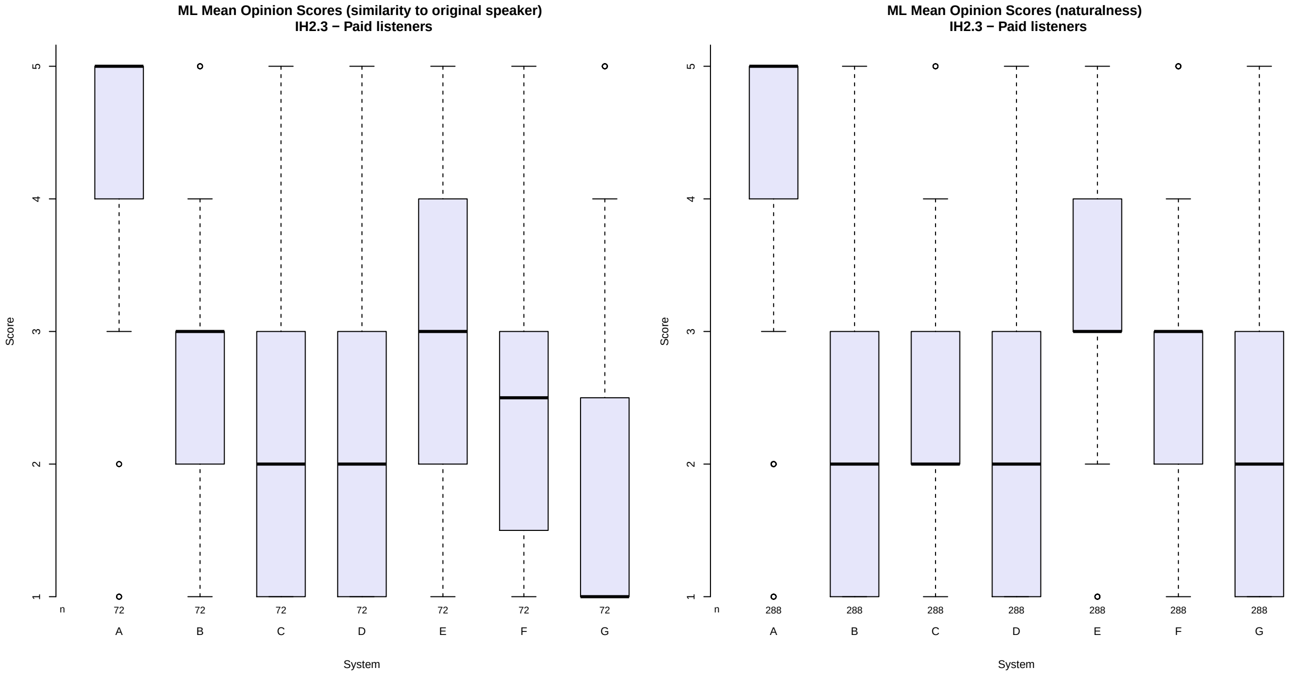

Figure 21: Similarity and Naturalness results on ML for IH2.3 (Malayalam)

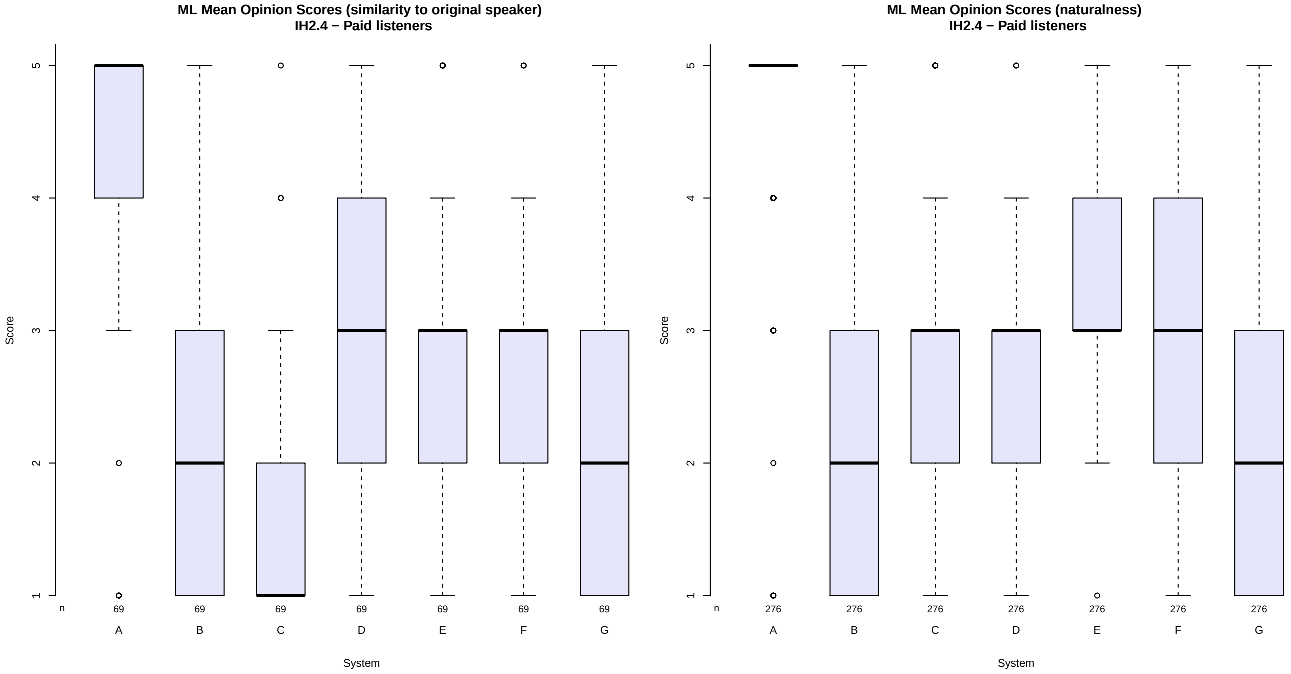

Figure 22: Similarity and Naturalness results on ML for IH2.4 (Marathi)

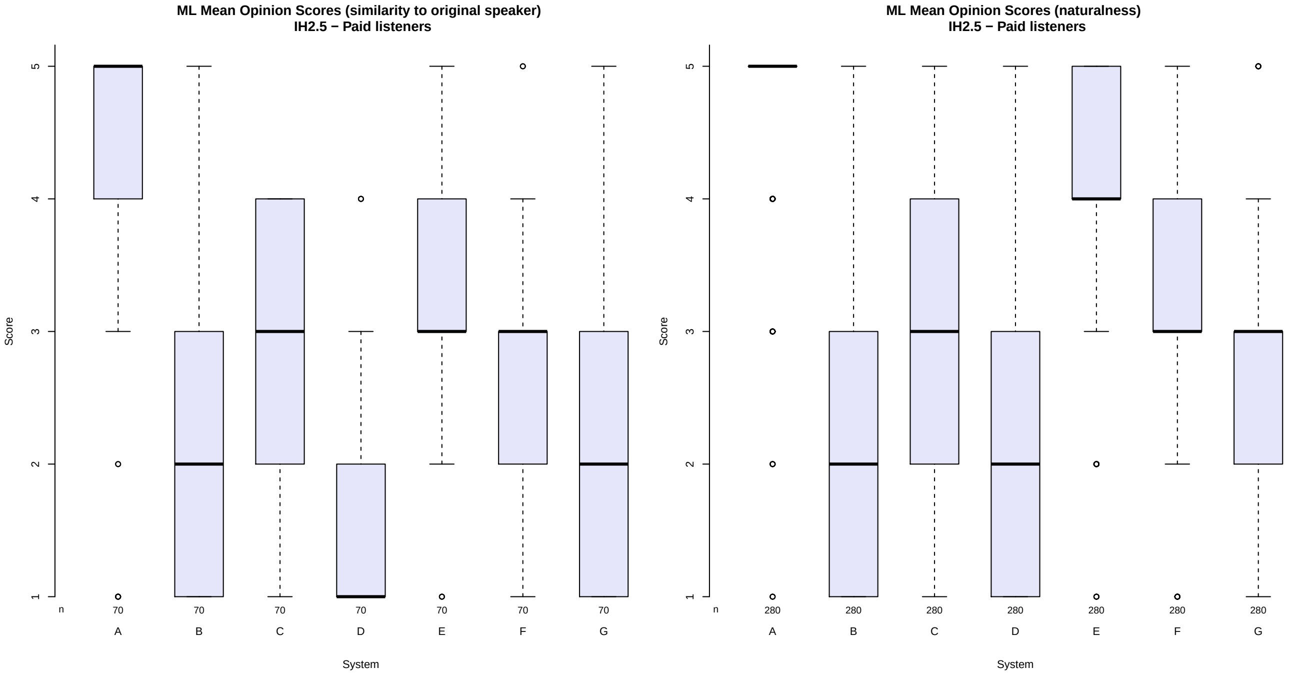

Figure 23: Similarity and Naturalness results on ML for IH2.5 (Tamil)

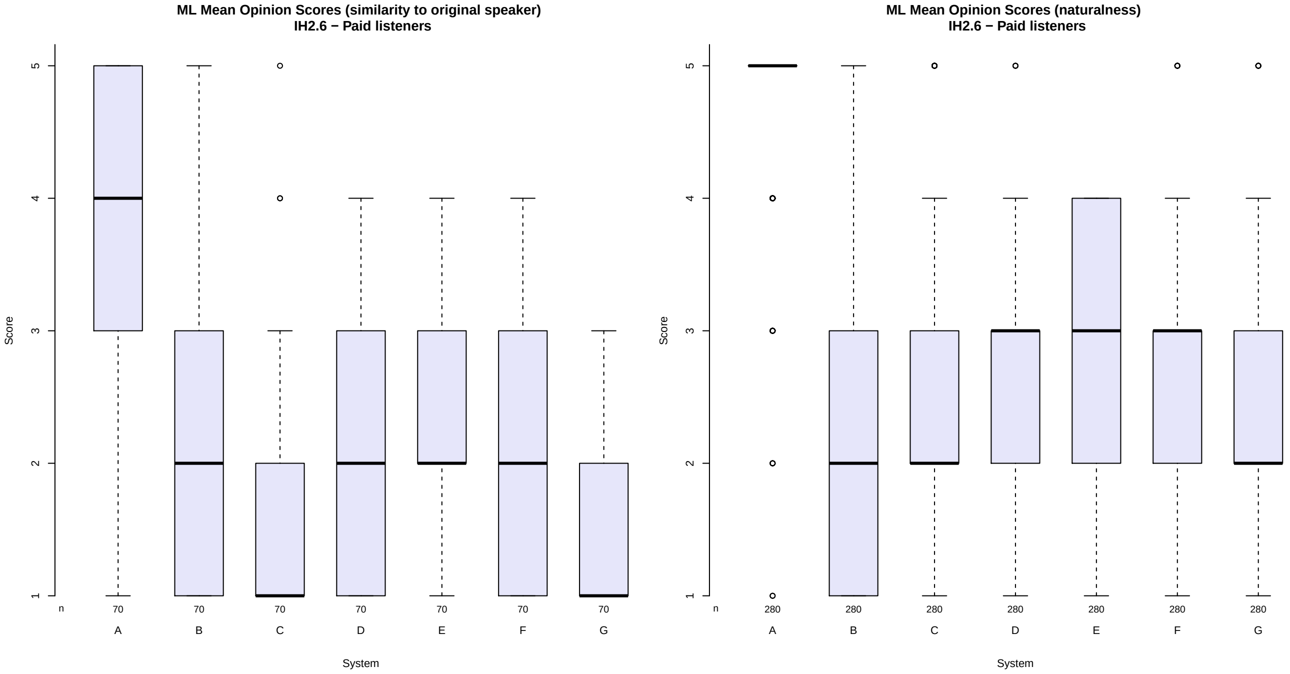

Figure 24: Similarity and Naturalness results on ML for IH2.6 (Telugu)