Human Interaction and Emerging Technologies (IHIET 2023), Vol. 70, 2023, 25–29
https://doi.org/10.54941/ahfe1002926

Human-Centered Design of Voice Communications: Gender Aspects

Jan Holub and Yann Kowalczuk

Czech Technical University in Prague, Faculty of Electrical Engineering, Prague, 166 27, Czech Republic

ABSTRACT

Potential misbalance of transmission quality between male and female speakers can affect many professionals that use distant voice communication in their daily duties. The subjective test laboratory has performed gender analysis for multiple subjective test projects run in 2019–2023. Not only traditional narrow-band voice coding principles but also some contemporary wide-band or even full-band digital communications show statistically significant differences between quality of transferred male and female voices. This article proposes to validate the gender balance in speech transmission quality subjective tests based on ITU-T P.800 and similar.

Keywords: Quality of experience, Human voice communication, Gender-balanced design, Subjective testing

INTRODUCTION

Perceiving the transmitted speech signal is a task that puts a certain amount of cognitive load on the human brain (Peelle 2018). The degree of this load depends on several factors, e.g., the loudness of the perceived speech, the type and intensity of background noise, the quality and accent of the speech, subject's familiarity with the topic of the message, etc. This load also varies between the native and non-native language (of the listener). Different levels of such load are manifested during longer task assignments / requirements (e.g., during a work shift) by different levels of overall fatigue, which affects the level of the worker's action or decision error rate when performing other concurrent tasks - the so-called parallel-task paradigm (Degeest et al. 2022).

For technologies used in speech transmission or synthesis, e.g., in telecommunications, radio communications, and machine to human communications, the above implies a strong need to optimize the coding of human (or synthetic) voice to minimize listening effort during communication (Marston 1998). Listening effort (LE) can be assessed by subjective tests following, e.g., ITU-T P.800 Recommendation, along with listening quality (LQ) specified in the same recommendation.

Gender Balance in Transmitted Speech Quality Testing – State of the Art

Often used methods of subjective testing of voice transmission quality are based on the procedures set out in ITU-T Recommendation P.800.

The principle is the evaluation of voice samples by many subjects and subsequent statistical evaluation (Grancharov and Kleijn 2008). The voice samples are created from original (undistorted) studio voice recordings either by real or simulated processing by the tested codec/technology/transmission channel. Usually, a number of groups of voice samples are prepared, called conditions, which differ e.g. in bit rate, background noise, combination with other contemporary codecs or jitter for packet transmissions, etc.

The resulting set of samples, assembled from each condition, is then played to test subjects in random order, and they evaluate either Listening Quality (LQ) or Listening Effort (LE) by means of subjective Opinion Score ranging from 1 (worst) to 5 (best). The original undistorted voice recordings (often labelled "c01") are also added to the sample set, as well as controlled distorted samples that have not been processed by the technology under test, which are used, for example, to compare results between different tests or even between laboratories that perform the tests. Final output parameter Mean Opinion Score (MOS) is then assigned to each tested condition as a mean value of the Opinion Scores assigned by all test subjects to the samples belonging to this condition. Other parameters like standard deviation and confidence interval (usually at $p = 0.05$) are calculated, too, to allow for comparisons between conditions and tests and for evaluating their statistical significance (usually performing paired or unpaired T-test or Z-test, depending on experiment and comparison type).

Gender equality is currently represented in this process by two requirements - an equal representation of male and female voice samples in the studio originals (namely, minimum two male and two female voices are required, each being represented by multiple sentences), and a requirement for an approximate balance between male and female test subjects when performing the subjective test, with a subsequent requirement to report this ratio in the final laboratory report.

As can be seen from the above description, the resulting quality of the male and female samples is not compared or even determined separately. Other test methods, e.g ITU-R BS.1534-3 "Multiple Stimuli with Hidden Reference and Anchor" (MUSHRA) follow similar approach to gender aspects, requiring balanced number of male and female input samples with no further gender-oriented analysis.

Problem Statement

A natural - but nowhere currently explicitly mentioned - requirement is that male and female voices are transferred with similar LQ and LE parameters; in other words, the transmission technology, including coding algorithms, frequency filters, or sampling rates, should not privilege one gender over the other to maintain similar working conditions and opportunities for all. Contrary to low interest in gender aspects in speech coding and transmission, closely related area of speech recognition deals with gender aspects with greater care (Oh et al. 2018, Ali et al. 2007). Potential gender coding/transmission quality misbalance can affect many professionals that deploy distant voice communication in their daily duties – e.g. female airport approach

Table 1. Comparison between AM and TETRA live transmission subjective quality (Kowalczuk 2022).

Sample group	Mean opinion score (MOS)	MOS_{Female}–MOS_{Male}	Statistically significant?	CI95
AM Female	2,297	**−0,192**	YES	0,052
AM Male	2,489			0,054
Tetra Female	3,523	**0,089**	NO	0,072
Tetra Male	3,434			0,069

control dispatchers or other professionals (policewomen) are principally handicapped by technological aspects of their job - worse voice transmission quality means higher listening effort is needed and may lead to consequent (subconscious) discomfort of their communication partners. Of course, gender transmission quality misbalance is not surprising for narrow-band or even old analog AM transmissions (as still used in AIRCOM) due to the generally higher pitch region of female voices (and can only be used as an argument to upgrade such communication means to a suitable digital format there), see Table 1.

Work Performed & Results Analysis

Additional gender analysis has been evaluated on multiple subjective test projects that subjective testing laboratory Mesaqin.com Ltd. performed since 2019 to see how (mis-)balanced the transmission/coding quality between male and female speakers is. The authors have all the detailed data of the described experiments, including precise descriptions of technologies and conditions tested, but due to existing contractual agreements the experiments are presented anonymously, only under their experiment number. However, the authors believe that despite the necessary anonymization, the results illustrate well the state of the art in contemporary technologies for voice coding or transmission.

Apart from the overall gender analysis (MOS calculated across all subjective scores assigned to samples derived from male voices versus MOS calculated across all subjective scores assigned to samples derived from female voices), also similar analysis of reference (undistorted) condition "c01" has been performed. The overall results of the gender analysis of reference samples are shown in Table 2.

Four out of 13 totally analysed tests (tests 1, 2, 10 and 11) were tests of narrow-band technology, limiting frequency response of the communication channel to the interval app. 300–3400 Hz. Two of those NB test showed male voices being transmitted subjectively better while the other two showed inconclusive results due to already misbalanced reference samples.

All other 9 tests were either in wide-band (WB, 50–7000 Hz), super-wide-band (SWB, 50–14000Hz) and one in full-band (FB, 20–20000Hz). The results indicate that some contemporary wide-band or even full-band digital communications also show statistically significant differences between the quality of coded/transferred male and female voices. In particular, seven

Table 2. Gender analysis result for experiments performed between 2019–2023.

Exp.	Test language	Band-width	MOS_{Fem}–MOS_{Male}	Stat. signif.?	Ref. sample analysis	Result
1	English	NB	0,50	M better	M ref better	unclear
2	English	NB	0,34	M better	M ref better	unclear
3	English	WB	0,39	M better	Ref OK	NOT BALANCED
4	English	WB	0,20	M better	Ref OK	NOT BALANCED
5	English	SWB	0,09	M better	Ref OK	NOT BALANCED
6	English	SWB	0,19	M better	Ref OK	NOT BALANCED
7	English	FB	0,30	M better	Ref OK	NOT BALANCED
8	Mandarin	FB	0,02	OK	Ref OK	balanced
9	English	NB	0,31	M better	Ref OK	NOT BALANCED
10	English	NB	0,47	M better	Ref OK	NOT BALANCED
11	English	NB	0,19	M better	Ref OK	NOT BALANCED
12	English	WB	0,25	M better	Ref OK	NOT BALANCED
13	English	WB	0,21	M better	M ref better	unclear

of those tests revealed statistically significant better transmission quality of male voices, one showed inconclusive results due to already misbalanced reference sample set and one condition showed gender-balanced results. It is interesting to note that this single gender-balanced test was performed in Mandarin Chinese language while all other reported tests were in English.

The language dependence of the results of this gender balance study is beyond the scope of this paper and remains a subject for further study, possibly using more experiments from multiple laboratories. The following observations can be derived from the achieved results:

Minimum frequency response of the future coders and transmission bandwidth in general should be maintained wide enough to preserve details of female higher-pitched voices in a similar manner as of male voices. Similarly wide frequency response should be exercises also by all other signal processing elements and algorithms (e.g., echo cancellation, noise reduction, automatic gain control algorithms) and electro-acoustic components including terminals (microphone/loudspeaker). However, our results show that enough wide bandwidth does not automatically guarantee gender balance between various voices, as other aspects affect the balance, too, e.g. the granularity of the perceptual frequency scaling used in codecs (Bark scale among others) or speech coding dynamics (e.g., voiced packet length, reconstructed signal smoothing).

CONCLUSION

In this article, we addressed the issue of speaker gender effect on transmission quality. The results confirm the necessity to ensure that future technologies for human voice coding and transmission aim to equalize transmission quality between the genders. It means their future subjective tests should result in statistically insignificant gender differences. Also, the use of qualitatively gender-balanced reference samples to enable proper consequent gender analysis of the entire subjective test should be preferred.

ACKNOWLEDGMENT

The authors would like to acknowledge subjective testing laboratory Mesaqin.com Ltd. for providing the subjective test datasets for our analysis.

REFERENCES

Ali, S., Siddiqui, K., Safdar, N., Juluru, K., Kim, W., Siegel, E. (2007), "Affect of Gender on Speech Recognition Accuracy", in: American Journal of Roentgenology, ISSN 0361-803X, American Roentgen Ray Society

Degeest, S., Kestens, K., Keppler, H. (2022), "Listening Effort Measured Using a Dual-task Paradigm in Adults With Different Amounts of Noise Exposure", in: Ear and Hearing 43(3): p 899-912, ISSN 0196-0202, DOI: 10.1097/AUD.0000000000001138, Lipincott Williams & Wilkins

Grancharov, V., Kleijn, W. (2008), "Speech Quality Assessment", in: Benesty, J., Sondhi, M., Huang, Y. (eds.) Handbook of Speech Processing, pp. 83–99. Springer

Kowalczuk, Y., Holub, J. (2022), "Can Gender Analysis Improve Quality?", in: ETSI STQ Workshop - Quality of Emerging Services for Speech and Audio: A user-centred perspective, Bratislava, SK

Marston, D.F. (1998), "Gender adapted speech coding", in: Proceedings of the 1998 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), IEEE Signal Processing Society

Oh, E., Yoo, H (2018), "Gender recognition using compressed speech data", in: Basic & Clinical Pharmacology & Toxicology, ISSN 1742-7835, Wiley

Peelle, Jonathan E. (2018) "Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior", in: Ear and Hearing 39(2): p 204-214, ISSN 0196-0202, DOI: 10.1097/AUD.0000000000000494, Lipincott Williams & Wilkins