Characterization of Human Emotions and Preferences for Text-to-Speech Systems Using Multimodal Neuroimaging Methods

Khalil ur Rehman Laghari*, Rishabh Gupta*, Sebastian Arndt †, Jan-Niklas Antons†, Sebastian Möller†
and Tiago H. Falk*
*INRS-EMT, University of Quebec, Canada †Quality and Usability Lab, Berlin Institute of Technology, Germany

***Abstract*—Voice user interface and speech quality are normally assessed using subjective user experience testing methods and/or objective instrumental techniques. However, the recent advances in neurophysiological tools allowed useful human behavioral constructs to be measured in real-time, such as human emotion, perception, preferences and task performance. Electroencephalography (EEG), and functional near-infrared spectroscopy (fNIRS) are well received neuroimaging tools and they are being used in variety of different domains such as health science, neuromarketing, user experience (UX) research and multimedia quality of experience (QoE) discipline. Therefore, this paper describes the impact of natural and text-to-speech (TTS) signals on a user's affective state (valence and arousal) and their preferences using neuroimaging tools (EEG and fNIRS) and subjective user study. The EEG results showed that the natural and high quality TTS speech generate "positive valence", that was inferred from a higher EEG asymmetric activation at frontal head region. fNIRS results showed the increased activation at Orbito-Frontal Cortex (OFC) region during decision making in favor of natural and high quality TTS speech signals. But natural and TTS signals have significantly different arousal levels.**

***Index Terms*—Text-to-Speech, TTS quality, User Experience, Emotions, Physiological metrics, EEG, fNIRS**

I. Introduction

Text-to-Speech (TTS) systems have seen a tremendous increase in recent years as emerging voice user interfaces (VUI) application for computers, smartphones, global positioning systems, and assistive technologies (e.g., for the visually impaired), to name a few. Thanks to advancements in technology, the intelligibility of TTS systems have improved multi-fold [1], however, their quality is still no match to natural human speech and this may cause a poor user listening experience.

The benchmarking of speech systems on "quality" is a cornerstone of all subjective and objective testing methods, however, there are also some other important human behavioral characteristics such as emotions, and user preferences, which should also be investigated in order to guage overall user experience. This paper aims to fill this gap by investigating user affective state and preferences with respect to natural speech and two different TTS speech quality signals.

In addition to subjective user testing and objective instrumental methods, another emerging trend is to use neuroimaging tools. These tools assist with detailed evaluation, allowing for behavioural insights to be gathered that would otherwise not be made available via questionnaires. For example, electroencephalography (EEG) has been used to obtain neural correlates of perceptual [2], [3], affective [4], [5], and cognitive processes [6]. Similarly, functional near-infrared spectroscopy (fNIRS), has also been used to understand human decision making process [7], speech perception [8] and emotions [9].

The human brain is complex organ that controls everything we do. The frontal brain is normally responsible for emotion processing [10]. While orbito-frontal cortex (forehead) is responsible for valuation based decision making [11]. In current work, EEG electrodes placed at frontal brain were investigated in order to evaluate affective state (valence and arousal). While fNIRS optodes were placed at forehead to measure neuronal activation in order to infer the user preference based decision making. Forehead, due to its proximity to eyes and facial muscles, is more susceptible to eye blinks, eye movements, facial muscles and other forehead artefact. Therefore, EEG electrode placement at forehead region is normally not recommended because it has more risk of getting contaminated with these artefacts than fNIRS [12]. Secondly, hair attenuate fNIRS optical signals, since forehead is free from hair, fNIRS becomes right choice to investigate this region of the brain.

The remainder of this paper is organized as follows. Section 2 provides an overview on TTS speech quality testing methods and neuroimaging tools and features, section 3, provides experimental details of current study. Section 4 provides subjective user study and neurophysiological results. Finally in section 5, we discuss and conclude our work.

II. Background

TTS speech quality is normally assessed either subjectively or objectively. Subjective user testing typically involves user ratings and surveys to collect speech quality of experience metrics. ITU-T P.85 is well known standard for subjective quality testing for TTS systems [13]. Objective methods, on the other hand, replaces the users with a computational algorithm that has learned complex mappings between several key factors and previously-recorded user ratings. Some of the objective methods for TTS quality assessment are discussed in [14], [15]. Another emerging trend is to use physiological tools to characterize speech quality of experience. It is known that human emotions and perceptions are outcomes of complex multimodal processing at the neuronal level. To measure emotions and other important behavioral characteristics, neuroimaging tools can be used. EEG is one such emerging tool, and it measures the electric potential of underlying neurons and it is recorded by placing electrodes on the head. Another tool is fNIRS and it measures the neuronal activation of the brain by measuring blood flow. For more detail on these techniques, interested readers can refer to [16].

To characterize human affective states, the human frontal head region has widely been investigated. Seminal studies have shown differential involvement of right and left hemispheres in emotional processing, where the activity in the right hemisphere is linked with unpleasant emotions and the left with pleasant emotions [17], [18]. As such, an asymmetry index feature has been developed which measures the difference in EEG activity in the alpha band from the right to the left hemisphere; the index is correlated with the subjective valence [10], [19]. Moreover, the beta frequency band power at the medial prefrontal cortex (MPC) have been linked to the arousal. Cortical activity at MPC can be investigated using the event-related synchronization or desynchronization (ERS/ERD) method. The higher the event related desynchronization in beta band, the lower will be the cortical activity.

fNIRS features are based on changes in blood flow, representing activation/deactivation of a specific brain region. The higher value of oxygenated ($\Delta[HbO]$) peak amplitudes and the lower values of deoxygenated haemoglobin ($\Delta[HbR]$) valleys indicate the increased activation. According to the so-called neuro-economics literature, the OFC is also responsible for the valuation and outcome evaluation processes involved in decision making [11], [20], [21]. As reported in [22], [23], the activation of the peaks and the valleys of the oxygenated ($\Delta[HbO]$) and deoxygenated haemoglobin ($\Delta[HbR]$) at OFC are linked with valuation based decision making process.

III. Experimental Methodology

A. Subjects

Fourteen fluent English speakers (6 Males, 8 Females) with an average age of 21.6 years participated in the subjective listening test. None of them reported having any hearing impairments or other health issues. In-ear headphones were used to play the synthesized speech stimuli at their individual preferred volume. The protocol was approved by the INRS and McGill Research Ethics Offices and participants consented to participate and were compensated monetarily for their time.

B. Synthesized speech stimuli

Synthesized speech stimuli used in this study were taken from the 2009 Blizzard TTS Challenge data [24]. The Challenge was developed to compare existing corpus-based TTS systems on the same development set.

The two TTS signals were selected and they were representative of low- (LQ), and high-quality (HQ) systems. For bench marking purposes, we also used the original "Natural" speech signal. The each stimulus comprised four English sentences (neutral in content) of duration 8-10 seconds. All stimuli were presented to the listeners at a sampling rate of 16 kHz and a bitrate of 256 kbps. The speech signals correspond to restaurant recommendation system.

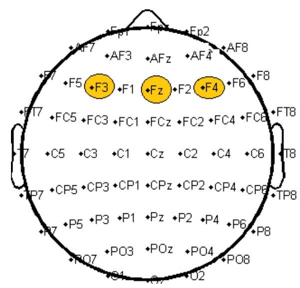

Fig. 1: EEG electrode positions on head as per 10-20 system, the highlighted electrodes are investigated in current paper

C. Experimental Protocol

At first, subjective ratings were collected for all 12 stimuli (4 sentences and 3 conditions- natural, HQ and LQ TTS). The self assessment manikin (SAM) pictorial system was used to measure valence and arousal, which ranges from "1" to "9", where 1 represents low valence/arousal and 9 represents the highest valence/arousal state [25].

After subjective user testing, participants were first fitted with EEG cap and customized fNIRS headband and then placed in front of a 22-inch computer screen. They were given a keyboard to provide binary response as whether they liked a stimulus presented to them or not. The ratio of stimulus with favorable rating to the total number of presented trials were calculated and this ratio factor was named as *Preference metric*. This part of the experiment was divided into six blocks, each lasting for about 10 minutes with an inter-stimulus interval (ISI) of around 20s. This ISI duration is important for fNIRS signal activity normalization, because it gives enough resting time for the changes in cerebral hemodynamics to return to baseline levels. The stimuli were pseudo-randomized within blocks and subjects.

D. EEG

For EEG recording, we used a 64 channel Biosemi system, with electrodes arranged in the 10-20 standard system, four electrodes for electrooculography (EOG) and two for mastoid-electrodes (right and left) were also used for reference. Data were recorded at 512 Hz but down-sampled to 256 Hz and band-pass filtered between $1-50$ Hz for offline analysis.
All channels were re-referenced to the average of all EEG-channels. EEG epochs, time locked to the onset of the stimuli, were extracted and averaged separately for each speech quality conditions and for each participant. Body movement and eye-blink artefact from data were removed using visual inspection and independent component analysis (ICA).

For EEG analysis, EEGLAB toolbox for the MATLAB [26] was used. Two EEG features were extracted, namely an alpha-band asymmetry index (AI) and the MPC beta power (MBP), as correlates of valence and arousal emotion primitives, respectively. More specifically, the AI feature was computed by subtracting the natural logarithm of the alpha power of the two frontal electrodes F3 and F4 (exterior electrodes highlighted in Fig. 1), as suggested by [10]:

$$AI = \ln(\alpha_{F4}) - \ln(\alpha_{F3}). \tag{1}$$

TABLE I: Repeated measure ANOVA for subjective affective factors across three different speech quality conditions

Variables	**p**	**F- Value**	η^2
Arousal	≤ 0.01	16.30	0.56
Valence	≤ 0.01	33.88	0.72

The MBP feature, in turn, was computed using event related desynchornization (ERD) with the beta power in the Fz position, where reference represents the beta sub band power at Fz during resting state, and original signal represents the beta sub-band power at Fz during audio stimulus.

$$MBP = (\beta_{Fz(Ref)} - (\beta_{Fz(Org)}/\beta_{Fz(Ref)}) * 100. \quad (2)$$

E. NIRS Signal Acquisition and Analysis

The NIRScout system was used with a customized headband. It has two probing wavelengths 760 and 850 nm. It comprised of 5 transmitters and 9 detectors with a minimum of 2.5 cm and a maximum of 3.4 cm inter-optode distance, thus resulting in 21 channels as shown in the Fig. 2(a).

Recordings were made at a sampling frequency of 10.42 Hz. Note that channels 17-18 correspond to the OFC region. fNIRS data were preprocessed and analyzed using the fNIRS-SPM toolbox [27]. The raw intensity signals from each channel were detrended using a discrete cosine transform based algorithm and converted into concentration levels of oxygenated ($\Delta[HbO]$) and deoxygenated haemoglobin ($\Delta[HbR]$) using the well-known modified Beer-Lambert law (MBLL) [28].

Two features were extracted from the two detrended waveforms, as described in [23]. The features included: peak amplitude of the $\Delta[HbO]$ curve and the amplitude of the $\Delta[HbR]$ curve valley. These features were extracted from each of the 21 functional channels for each participant. The $\Delta[HbO]$ peak, abbreviated as "OP" and $\Delta[HbR]$ valley, abbreviated as "DV" respectively. They have been found to be correlated with the Blood Oxygenation Level Dependent (BOLD) signal measured via magnetic resonance imaging, which in turn is positively correlated with regional neural activation [29].

IV. Results

A. Characterization of Affective State

Affective states (valence, arousal) were collected using subjective SAM scale across three difference speech quality conditions. A repeated measure within subjects ANOVA was computed on affective data using the predictive analytic software SPSS. The effects of three different quality conditions (natural, HQ, LQ TTS) were compared on each response variable: valence and arousal. Mauchly's test indicated that the assumption of sphericity had been violated ($p < 0.05$), therefore degrees of freedom were corrected using Greenhouse-Geiser estimate of sphericity. The ANOVA results are reported in Table I, and they show a significant main effect in subjective response variables across three different quality conditions. As can be seen, there is a strong effect size η^2 for valence but arousal has moderate effect size score.

TABLE II: Post-Hoc test based on paired-samples T-test for Arousal, and Valence

Conditions	**Arousal**	**Valence**
Nat-HQ	$p \leq 0.01$	$p \leq 0.01$
Nat-LQ	$p \leq 0.01$	$p \leq 0.01$
HQ-LQ	$p = 0.5$	$p \leq 0.01$

TABLE III: Paired T-test for AI and MBP with Cohen's D effect size

Conditions	**AI**	**Effect Size**	**MBP**	**Effect Size**
Nat-HQ	$p = 0.80$	0.07	$p \leq 0.02$	1.10
Nat-LQ	$p \leq 0.04$	0.73	$p \leq 0.05$	0.87
HQ-LQ	$p \leq 0.05$	0.68	$p = 0.47$	0.28

As a post-hoc analysis, paired T-test was computed as given in Table II. Valence score passed paired-samples T-test across three different conditions. However arousal did not show significant difference between TTS condition (HQ-LQ). This finding indicates that subjects did not feel difference in arousal level between HQ and LQ TTS signals.

The subjective findings reported above suggest that perceived quality and induced affect play a crucial role in user quality of experience. As such, in order to develop a true objective metric, EEG based neural metrics are needed. Asymmetry Index (AI) is widely known as counter part of subjective valence score [10]. It was computed to evaluate valence levels. For the measurement of arousal level, MBP was computed using event related desynchronization method (ERD), as given in equation 2. MBP is normally used to measure the arousal characteristics (e.g., attention, engagement).

After preprocessing of EEG data, hypothesis testing on AI and MBP features across three different quality conditions was conducted. And effect size was computed using Cohen's D method. Results are depicted by Table III. Unlike subjective results, a paired-samples T-test on AI scores indicated that natural and HQ TTS signal scores were not significantly different, and their low effect size also confirmed the same (D=0.07). However, paired-samples T-test between (Nat-LQ) and (HQ-LQ) indicated that these signals are significantly different, and Cohen's D values also indicated the strong effect size. These results suggest that natural and HQ TTS evoked increased asymmetric activation or similar lateralization effects on prefrontal brain in contrast to LQ TTS signal. This is probably because natural and HQ TTS have more pleasant quality than LQ TTS speech signals.

Paired-samples T-test on MBP scores was also computed and the (Nat-HQ) and (Nat-LQ) conditional pairs showed significant mean differences, and Cohen's D values also indicated the strong effect size as shown in Table III. However, the test failed for HQ-LQ TTS condition. It means, on the basis of MBP data, one can infer that natural speech has significantly different mean values than HQ and LQ TTS signal. This is probably because natural speech has different prosodic nature than TTS signals, and in this experiment, natural speech signal was also deviant signal between the pool of TTS signals, and every time, a subject listed to natural signal, it might have

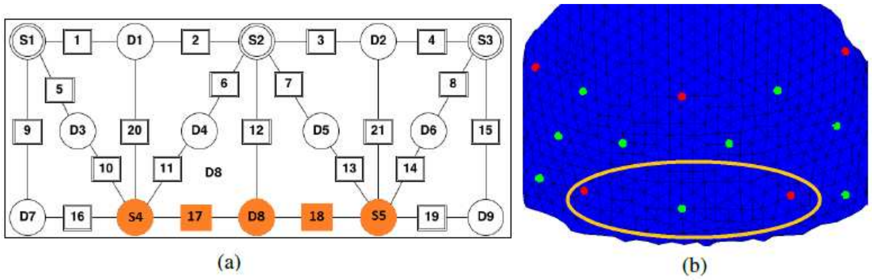

Fig. 2: fNIRS headband optode topology where 2 (a) shows the 21 channels, depicted within squares with ‘S’ and ‘D’ showing the source, the detector positions, colored area is area of investigation and 2 (b) presents the 3-D finite element method (FEM) head model with the source and detector shown in red and green, respectively.

TABLE IV: Repeated measure ANOVA for fNIRS features

Features	p	F-Value	η^2
Ch17OP	$\leq$ 0.02	4.77	0.35
Ch17DV	$\leq$ 0.05	7.37	0.45
Ch18OP	$\leq$ 0.07	2.91	0.23
Ch18DV	$\leq$ 0.02	4.40	0.30

evoked more attention and engagement. These results were also corroborated by the subjective arousal.

Next, the Pearson and Spearman rank correlation values were also computed between subjective data and EEG affective features as presented in Table VI. AI showed a significant positive correlation with the subjective valence ratings (0.44; $p \leq 0.05$), thus, indicating a similar increasing trend between the AI metric and subjective valence score. The MBP ERD metric, in turn, showed a significant negative trend with subjective arousal ratings (-0.44; $p \leq 0.05$); this is because MBP ERD is inversely related to arousal and excitement.

B. Characterization of Preference

If we humans consider a task more rewarding, we may prefer to do it, this value based decision making activates OFC region as discussed in [11]. We used fNIRS headband on forehead to capture OFC signals. The fNIRS features specially the amplitudes of the $\Delta[HbO]$ peaks and $\Delta[HbR]$ valleys were computed. The amplitudes of the $\Delta[HbO]$ peaks increase and $\Delta[HbR]$ valleys decrease with higher perceptual quality in the OFC region (channels 17 and 18), suggesting increased activation of the region during decision making.

To provide statistical evidence, a within subjects repeated measures ANOVA was used. Mauchly’s test indicated that the assumption of sphericity had been met($p > 0.05$) for the fNIRS features. A significant main effect was observed for fNIRS features (except for the channel 18 $\Delta[HbO]$ peaks) across all three quality conditions as reported in Table IV, but the effect size was found to be weak for all fNIRS features.

For further investigation, Post-hoc analysis based on paired-samples T test was computed. As shown in Table V, there was no significant mean difference between natural and HQ TTS and while for all other conditions, a significant mean difference was found. The higher values of $\Delta[HbO]$ peaks and lower values of $\Delta[HbR]$ valleys for natural speech and HQ

TABLE V: Post-Hoc test based on Paired T-test for fNIRS features

Conditions	CH17OP	CH17DV	CH18OP	Ch18DV
Nat-HQ	p =0.40	p =0.90	p =0.90	p =0.40
Nat-LQ	p $\leq$0.01	p $\leq$0.01	p $\leq$0.06	p $\leq$0.01
HQ-LQ	p $\leq$0.01	p $\leq$0.01	p $\leq$0.01	p $\leq$0.05

TABLE VI: Correlation Matrix between Neurophysiological features and Subjective factors

Variables	Pearson Correlation	Spearman Correlation
Valence- AI	0.44	0.42
Arousal-MBP (ERD)	-0.44	-0.43
Pref-CH17OP	0.49	0.50
Pref-CH17DV	-0.63	-0.66
Pref-CH18OP	0.33	0.24
Pref-CH18DV	-0.52	-0.51

TTS signal in comparison to LQ TTS signal suggest that the former stimuli induce a significantly higher activation at OFC region (channel 17 and 18). These channels were located on OFC region as shown in Figure 2. The higher activation of the OFC region indicates the higher valuation assigned to natural and HQ TTS signals. This finding is further corroborated by the well correlated Preference metric and fNIRS features ($\Delta[HbO]$ peaks and $\Delta[HbR]$ valleys) from channels 17 and 18 as shown in Table VI.

V. Discussion and Conclusion

In this paper, we attempted to characterize TTS speech signals based on user emotions and preferences using subjective user ratings, and neuroimaging tools.

Our findings based on subjective user ratings for TTS system evaluation using “valence” measure suggest a linear trend between speech quality and subjective valence. Repeated measure ANOVA and Post-Hoc tests validated this trend. However, EEG based valence measure, Asymmetry Index has slightly different results, for instance, paired T-test failed between natural and HQ TTS signals, suggesting that both signals demonstrated similar asymmetric lateralization or activation, unlike LQ TTS. fNIRS data for preference based decision making (i.e., activation at OFC) region also followed the similar trends as of EEG AI trends. It means, natural and

HQ TTS signal has similar valence and user acceptance levels, except LQ TTS signal. However subjective valence ratings might have been influenced by other factors (e.g., arousal).

Interestingly, arousal factor followed a different trend, unlike valence and preference metric. Our findings based on subjective user ratings and EEG metric suggest that natural speech signal has different level of arousal (excitability level) than HQ and LQ TTS signal (see Table II, III), while HQ TTS signal did not demonstrate a significant difference in arousal level with LQ TTS signal. Natural speech signal surpassed both TTS signals on the basis of "arousal" level. Thus, it also signals the need for improvement in TTS signals in terms of arousal factor, despite their natural sounding prosody and high quality delivery, they are still lagging behind the natural speech signal. It is also possible that by excluding natural signal from study and re-testing only TTS signals may generate different arousal levels. However, it is not recommended, as natural sounding quality and feel is the ultimate target of TTS systems.

TTS system developers should test their system not only on "quality" benchmarks, but also on emotions and other cognitive characteristics. In order to obtain subtle details and indepth view on overall user experience with TTS system, physiological tools become good addition in testing toolkit.

Acknowledgment

THF is grateful to the Canadian Foundation for Innovation and the Ministere de l'Education, du Loisir et du Sport du Quebec for funding the experimental facility and the Ministere du Developpement Economique, Innovation et Exportation du Quebec and the National Science and Engineering Research Council of Canada for funding the pilot studies. The co-authors from Germany acknowledge funding from the Federal Ministry of Education and Research (Grant FKZ 01GQ0850).

References

[1] K. Parssine, "Multiligual text-to-speech system for mobile devices: Developement and applications," Ph.D. dissertation, University of Technology Tampere, 2007.

[2] C. Duncan, R. Barry, J. Connolly, C. Fischer, P. Michie, R. Naatanen, J. Polich, I. Reinvang, and C. Van Petten, "Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400," *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1883 – 1908, 2009.

[3] K. Laghari, R. Gupta, S. Arndt, J. Antons, R. Schleicher, S. Moller, and T. Falk, "Auditory BCIs for Visually Impaired Users: Should Developers Worry About the Quality of Text-to-Speech Readers?" in *International BCI Meeting*, 2013.

[4] J. Antons, R. Schleicher, S. Arndt, S. Moller, A. Porbadnigk, and G. Curio, "Analyzing speech suality serception using Electroencephalography," *IEEE J. Select. Topics Signal Proc*, vol. 6, no. 6, pp. 721–731, 2012.

[5] D. Bos. (2006) EEG-based emotion recognition: The influence of visual and auditory stimuli. [Online]. Available: http://hmi.ewi.utwente.nl/verslagen/capita-selecta/CS-Oude_Bos-Danny.pdf

[6] M. Rugg and M. Coles, *Electrophysiology of the Mind: Mental chronometry and the study of human information processing.*, M. Rugg and M. Coles, Eds. Oxford University Press New York, 1995.

[7] Y. Mitsuda, K. Goto, T. Misawa, and T. Shimokawa, "Prefrontal cortex activation during evaluation of product price: A NIRS study," in *Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference*, 2012.

[8] S. Yoo and K. Lee, "Articulation-based sound perception in verbal repetition: a functional NIRS study," *Frontiers in Human Neuroscience*, vol. 7:540, 2013.

[9] H. Doi, S. Nishitani, and K. Shinohara, "NIRS as a tool for assaying emotional function in the prefrontal cortex," *Front Hum Neurosci.*, vol. 7:770, 2012.

[10] J. A. Coan and J. J. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biological Psychology*, vol. 67, p. 749, 2004.

[11] A. Rangel, C. Camerer, and P. Montague, "A framework for studying the neurobiology of value-based decision making," *Nature Reviews Neuroscience*, vol. 9, no. 7, pp. 545–556, 2008.

[12] L. Gar Wai Ko, "Near-infrared Spectroscopy as an Access Channel: Prefrontal Cortex Inhibition During an Auditory Go-No-Go Task," Master's thesis, Institute of Biomaterials and Biomedical Engineering University of Toronto, 2008.

[13] F. Hinterleitner, G. Neitzel, S. Moller, and C. Norrenbrock, "An evaluation protocol for the subjective assessment of text-to-speech in audiobook reading tasks," in *Proc. Blizzard Challenge Workshop. International Speech Communication Association (ISCA)*, 2011.

[14] T. Falk and S. Moller, "Towards signal-based instrumental quality diagnosis for text-to-speech systems," *IEEE SIGNAL PROCESSING LETTERS*, vol. 15, pp. 781–784, 2008.

[15] I. Kraljevski, S. Chungurski, I. Stojanovic, and S. Arsenovski, "Synthesized Speech Quality Evaluation Using ITU-T P.563," in *18th Telecommunications forum TELFOR*, 2010.

[16] K. Laghari, R. Gupta, S. Arndt, J. Antons, R. Schleicher, S. Moller, and T. Falk, "Neurophysiological experimental facility for Quality of Experience (QoE) assessment," in *Proceedings of International Conference on Quality of Experience Centric Management (QCMan)*, 2013.

[17] L. J. Heller W, "Perception and expression of emotion in right-handers and left-handers." *Neuropsychologia.*, vol. 19(2), pp. 263–72., 1981.

[18] R. Davidson, *Hemispheric specialization for cognition and affect*, A. Gale and J. Edwards, Eds. Academic Press London, 1983.

[19] S. Arndt, J. Antons, K. Laghari, R. Gupta, R. Schleicher, S. Moller, and T. Falk, "The effects of text-to-speech system quality on emotional states and frontal alpha band power," in *IEEE EMBS Neural Engineering Conference*, 2013.

[20] A. Blood and R. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, pp. 11 818–11 823, 2001.

[21] S. Tom, C. Fox, C. Trepel, and R. Poldrack, "The neural basis of loss aversion in decision-making under risk," *Science*, vol. 315, no. 5811, pp. 515–518, 2007.

[22] E. Rolls and F. Grabenhorst, "The orbitofrontal cortex and beyond: from affect to decision-making," *Progress in Neurobiology*, vol. 86, no. 3, pp. 216–244, 2008.

[23] R. Gupta, K. Laghari, S. Arndt, R. Schleicher, S. Moller, and T. Falk, "Using fNIRS to Characterize Human Perception of TTS System Quality, Comprehension, and Fluency: Preliminary Findings," in *Proceedings of the 4th International Workshop on Perceptual Quality of Systems PQS 2013*, 2013.

[24] K. Simon and K. Vasilis, "The Blizzard Challenge 2009," in *Proceedings of Blizzard Challenge Workshop*, 2009.

[25] J. Morris, "Observations: SAM:The self assessment manikin, An effecient cross-cultural measurement of emotional response." University of Florida.

[26] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.

[27] J. Ye, S. Tak, K. Jang, J. Jung, and J. Jang, "NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy," *Neuroimage*, vol. 44, no. 2, pp. 428–447, 2009.

[28] D. Boas, T. Gaudette, G. Strangman, X. Cheng, J. Marota, and J. Mandeville, "The accuracy of Near Infrared Spectroscopy and imaging during focal changes in cerebral hemodynamics," *Neuroimage*, vol. 13, no. 1, pp. 76–90, 2001.

[29] N. Logothetis, "The underpinnings of the BOLD functional Magnetic Resonance Imaging signal," *The Journal of Neuroscience*, vol. 23, no. 10, pp. 3963–3971, 2003.