Subjective evaluation of olfactory and visual media synchronization — **Source link** ⧉

Niall Murray, Yuansong Qiao, Brian Lee, A. K. Karunakar ...+1 more authors

Institutions: Athlone Institute of Technology, Manipal University, Dublin City University

Related papers:

- Mulsemedia: State of the Art, Perspectives, and Challenges

- The sweet smell of success: Enhancing multimedia applications with olfaction

- Synchronization of Olfaction-Enhanced Multimedia

- Multiple-Scent Enhanced Multimedia Synchronization

- User Quality of Experience of Mulsemedia Applications

Subjective Evaluation of Olfactory and Visual Media Synchronization

Niall Murray[1], Yuansong Qiao[1], Brian Lee[1], Karunakar A. K.[2], Gabriel-Miro Muntean[3]

[1]Software Research Institute, Athlone Institute of Technology, Ireland
[2]Manipal Institute of Technology, Manipal University, India
[3]School of Electronic Engineering, Dublin City University, Ireland

{nmurray, ysqiao}@research.ait.ie, blee@ait.ie, karunakar.ak@manipal.edu, munteang@eeng.dcu.ie

ABSTRACT

As a step towards enhancing users' perceived multimedia quality levels beyond the level offered by the classic audiovisual systems, the authors present the results of an experimental study which looked at user's perception of inter-stream synchronization between olfactory data (scent) and video (without relevant audio). The impact on user's quality of experience (by considering enjoyment, relevance and reality) comparing synchronous with asynchronous presentation of olfactory and video media is analyzed and discussed. The aim is to empirically define the temporal boundaries within which users perceive olfactory data and video to be synchronized. The key analysis compares the user detection and perception of synchronization error. State of the art works have investigated temporal boundaries for olfactory data with audiovisual media, but no works document the integration of olfactory data and video (with no related audio). The results of this work show that the temporal boundaries for olfactory and video only are significantly different from olfactory, video and audio. The authors conclude that the absence of contextual audio reduces considerably the acceptable temporal boundary between the scent and video. The results also indicate that olfaction before video is more noticeable to users than olfaction after video and that users are more tolerable of olfactory data after video rather than olfactory data before video. In addition the results show the presence of two main synchronization regions. This work is a step towards the definition of synchronization specifications for multimedia applications based on olfactory and video media.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems - *Human Factors*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems - *Artificial, augmented, and virtual realities*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - *Evaluation/methodology*

General Terms

Human Factors, Design, Experimentation.

Keywords

Olfaction, Multimedia Synchronization, Subjective Quality Assessment, Quality of Experience

1. INTRODUCTION

Multimedia systems have been characterized by the integration, combination, presentation, storage and communication of independent discrete and continuous media such as: text, animation, graphics, images, audio and video. Today the research community is extending this list with so-called new media like e-touch [1], e-taste [2] and e-smell [3]. The result is the emergence of multisensory communication and experience. The rationale of enhancing multimedia applications to stimulate more than audiovisual senses is to increase the user's Quality of Experience (QoE) [5]. With a significant demand already placed on the audiovisual senses, another avenue to increase QoE is through stimulation of the other senses. Olfaction is the sense of smell. Recently, scents have been used in multimedia, in particular with movies as it is assumed that presenting the scent according to the scenes would deepen the viewer's understanding and sense of reality [6]. In addition, we can now find the use of olfaction across other industries in the literature; gaming [7], health [8], education [9], training [10], tourism [11] and entertainment [12].

Research on modeling and analyzing the human perception of multimedia experiences is an active topic [3][5][18][19][20][21]. It is now accepted that objective measures alone do not reflect the end user perception of a multimedia experience. This focus on the user is generally referred to as Universal Multimedia Experience (UME). It has been shown that humans perceive smell differently based on a number of factors, including age, culture, mood, gender and life experiences [22]. Interestingly, little research has been carried out in terms of perception of olfactory data with other media [3][4][16][23][24]. With olfaction, how the user perceives the experience is particularly important, considering the number of characteristics that affect its perception.

With respect to multimedia systems, synchronization is the process of determining and maintaining temporal relationships within and between media. Inter-stream synchronization involves the maintenance of temporal relationships between different streams e.g. audio and video. A difference in presentation times of related media data objects of different media is called the inter-stream "skew". Two media streams perfectly synchronized have no skew, i.e. 0 ms. Synchronization in multimedia systems is required when transmitting the above mentioned time dependent media streams (audio, video, haptic, olfactory etc.) across

communication networks as the temporal relationships between these streams can be disturbed because of network loss, delay and jitter, resulting in asynchrony.

The focus here is to determine, based on user's perspective, the temporal boundaries within which the level of inter-stream skew between olfactory data and video are viewed as being synchronized. In this paper, the use of the term video reflects the use of visual media only, the term audiovisual refers to the combination of audio and video. When compared with [16], where the relationship between olfactory and audiovisual media was studied, the results here show that the removal of contextual audio has a significant impact on user detection of skew, the scale of acceptable skew, as well as impacting reality, relevance and enjoyment. Cross-modal effects, i.e. the interaction of the senses, can have a major influence on how environments are perceived, even to the extent that large amounts of detail perceived by one sense may be ignored when in the presence of other more dominant sensory inputs [41]. Without contextual audio, the contribution of this work is to define the temporal relations between olfaction and video only using subjective studies. In terms of application areas, this work is applicable to on-line immersive gaming whereby the music or audio may not describe what is visually being presented, as also may be the case with TV (and which was the case with some of the videos used in [16]).

The remainder of this paper is organized as follows: Section 2 discusses related work, Section 3 presents challenges associated with the synchronization of olfactory and video media and Section 4 describes the components of the olfactory and video media display system used during the subjective testing. Section 5 outlines the assessment methodology employed, Section 6 presents the results and analysis of the completed subjective testing and Section 7 discusses our conclusions and directions for future research.

2. Related Works

A fundamental requirement of any multimedia application, including those enhanced with olfactory data, is synchronized display of multiple media. Research on synchronization of multi-sensory media applications is an active research area [15][25][26][27][3][4]. In the context of standardization, MPEG-V defines metadata representations for olfactory data among other sensory effects as part of its Sensory Effects Description Language (SEDL) within Sensory Information (part 3) [28].

In contrast to the traditional style of considering network delay and jitter for determining whether or not multiple media were synchronized, now the users' perception of the media experience is accepted as an important consideration. Little work has been documented on user perceived inter-stream synchronization of olfactory data with other media, with [16][29] for audiovisual and olfactory, [4] haptic and olfactory being the exceptions. The methodology used in these works was originally documented in [17][30]. In [17], inter-stream skews were artificially introduced between audio and video (lip synchronization) to determine the acceptable user perceived temporal synchronization boundary.

Works attempting to address the issue of scent lingering, approach this from the scent emitter perspective [33][34][35][36]. These works focus on the hardware that enables controlled emission of minute amounts of scent. The aim is to minimize scent lingering and "enable the instantaneous switching of scents" [35] through the precisely controlled presentation of olfactory data. It is arguable that these works are dealing with olfactory data from an intra-stream perspective. [24] reported a novel system for presenting movie with smell using a head mounted olfactory enhanced display. In addition they suggest a guideline relevant to the display of smell with audiovisual information such that "the period between certain smell and another different smell should be longer than 5s".

In [16], the methodology of [17] was employed to define a user perceived temporal boundary within which audiovisual data is synchronized with olfactory data. The authors introduced artificial inter-media skews between olfactory and audiovisual media and asked assessors to qualify their experience. In addition to defining audiovisual and olfactory temporal synchronization boundaries, they analyzed the impact of asynchrony in terms of annoyance, distraction, enjoyment, sense of reality and sense of relevance. The researchers found that olfaction before audiovisual content is more tolerable than olfaction after audiovisual content. Their work [16][29] is the closest to our work found in the literature. The key difference is that, while they focused on olfaction and audiovisual media, here the authors focus on the temporal relations between olfaction and video media only (audio is not related to video). In the next section, we introduce and discuss the olfactory and video media display system that was developed to enable this work.

3. Olfactory and Video Synchronization

Adding olfactory data as a media brings a number of challenges not common with text, graphics, audio or video media. Smell has a tendency to linger, it is slow moving media, "unlike the transitory nature of audio and video" [22]. In addition, it is important to recognize the existence of a number of phenomena associated with the olfactory sense. These are discussed in the following section.

3.1 Characteristics of Olfaction

Unlike video or audio, smell is a chemical media. Humans detect odors based on the interaction of odor molecules with smell receptors. Olfactory adaptation occurs when assessors are subjected to continuous olfactory stimulation. The sensory nerve activity decreases to a level where assessors find it difficult to perceive stimuli or don't perceive at all. With the removal of scents, perception is generally restored within a few minutes. Anosmia is another olfactory related phenomenon whereby there exists a lack of sensitivity to olfactory stimuli. It can be total, partial, permanent or temporary. It may result in an inability to perceive one or many different odors. Olfactory thresholds are values that express the amount of scent stimulus required to give an olfactory sensation. A number of sensory related thresholds are described in [37]. The detection threshold for any media is the minimum value of a sensory stimulus needed to give rise to a sensation without the sensation needing to be defined. The olfactory detection threshold "has strong appeal because it measures a feature of perception and performance in physical units of concentration" [38]. The recognition threshold is the minimum physical intensity of a stimulus for which an assessor will assign the same descriptor each time it is presented. The terminal threshold is the minimum value of an intense sensory stimulus above which, no difference in intensity can be perceived. In this work, the term detection instant is defined as the time at which assessors recognize the existence of an odor. This work analyses this instant in terms of the assessor perception of the synchronization between olfactory and video media.

4. Experimental Set-up

This section outlines the olfactory and video display, laboratory design, assessors as well as video and scents used in this work.

4.1 Olfactory and Visual Media Display

The following hardware and software were used for the subjective testing. As per Fig. 1, the olfactory and video display system consists of the ***SBi4 – radio v2 scent emitter*** (item Y sitting on laptop) from Exhalia [31]. It presents scents by blowing air (using 4 in-built fans) through scent cartridges. In version 2 of the SBi4, it is possible to control the intensity of the scent emitted by altering the fan speed. The greater the fan speed, the greater the intensity of scent presentation and the quicker it is presented to the assessors olfactory field. SBi4 can store up to four scent cartridges at any one time.
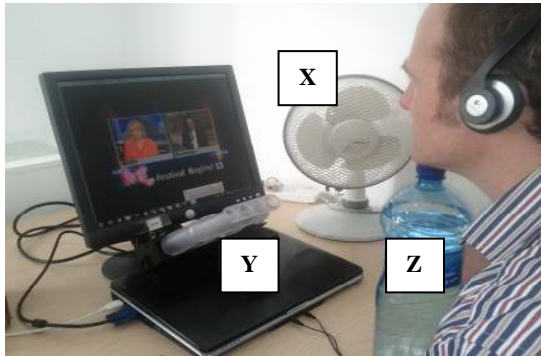

Fig. 1. Olfactory and Video media display system.

Fig. 2 shows the SBi4, scent cartridges and the bespoke extension that was designed and added to the SBi4 as shown in Fig 1 (Item Y). The purpose of this extension was to facilitate an accurate presentation of the scent to the users' olfactory field as opposed to a more general presentation. Based on the SBi4 being 0.5 meters from the assessor, it was found that it took assessors between 2.7s - 3.7s to detect the scents depending on the scent. Further discussion of how this was determined is documented in section 6.1. The cartridges of SBi4, exposed during operation, are made from scented polymer balls. Initially with the SBi4 cartridges, it is possible to detect odors in advance of any fans running (due to natural vaporization); however after 2-3 days, detectable odors are minimal and for most scents, not possible to detect at all when fans are not running.

The SBi4 system is controlled using the Exhalia java-based SDK. It is connected to the laptop via a USB port. The video content was played using ***the VLC media player 1.0.1 Goldeneye***. Presentation of the scent and video sequences was controlled using ***a special program that was developed to present both synchronized and skewed media components*** as per section 5.2. ***The laptop*** is windows 7 professional, Intel Core™ 2 Duo CPU @ 1.66GHz with 2GB RAM. ***The display screen was 21 inches*** with a resolution of 1024*768.

During the testing, assessors were seated at the testing booth shown in Fig. 1, in the experimentation room as shown in Figure 3. In addition, Fig. 1 also includes ***a bottle of water*** that the assessors placed under their chin during testing (Fig. 1, item Z). The purpose of this was to have consistency across all assessors in terms of the location of their olfactory fields regardless of posture or physical size. ***The fan*** in Fig 1. (Item X) is turned on between test sequences to remove any lingering scent.

Fig. 2. SBi4 V2, scent cartridges and bespoke extension.

4.2 Laboratory Design

The design of the test laboratory is in accordance with ISO standard [32], "Sensory analysis – General guidance for the design of test rooms". The aim of this standard is to design test rooms such that it is possible (1) to conduct sensory evaluations under known and controlled conditions with minimum distraction and (2) 4o reduce the effects that psychological factors and physical conditions can have on human judgment. The minimum requirement for the creation of test rooms are (1) a testing area in which work may be carried out individually in testing booths and (2) a preparation area. Fig. 3 shows a plan view of the preparation room and storage room for samples (room A), experimentation room (room B), and waiting room for assessors (room C). Walls in the test room are Matt off-white. One temporary testing booth (Fig. 3, B1) is situated in the corner of the test room to minimize distraction. Assessors complete questionnaires (see section 5) in the furthest point form the testing booth (B2). This allows time for scent to diffuse, minimizes adaptation, gives assessors a break between each judgement and avoids assessor being influenced by lingering scent in the air.
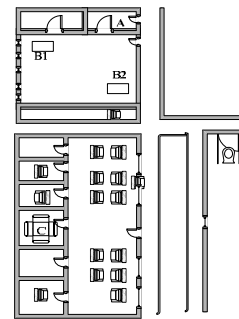

Fig. 3. Plan view of experimentation room (B), preparation room (A) and meeting room (C).

A sign restricting access to the test room is posted outside the door. The preparation area is located adjacent to the test room as shown. Assessors do not have access to this room. Finally, the office is where assessors have the opportunity to ask questions and sign necessary documentation in advance of any test is shown as meeting room C in Fig. 3. Whilst no ventilation system exists per se, the test lab is large, has 3 doors and in addition has multiple windows to remove scent from the test area. In addition, between viewing clips, the fan was turned on to remove lingering scent.

4.3 Assessors

A total of 43 assessors (20 female, 23 male) took part in the study. This group included people of multiple nationalities, between the ages of 19 to 56 from a wide variety of backgrounds: post

graduate researchers, academic staff, professionals and members of the public. In order to be eligible, assessors could not be involved in any sensory analysis testing in the twenty minutes preceding the tests. In an attempt to provide contamination free results, assessors must be free from colds or flu's; must avoid wearing perfume, aftershave or scented deodorants on the day of the testing. In addition they were requested to avoid chewing gum, eating food, drinking tea or coffee in the 30 minutes prior to the test.

4.4 Video Sequences and Scents

Six videos used (kindly provided by the authors of [16]) were of 90s duration (audio was removed). Each of the video clips can be divided into three 30 second blocks whereby the middle 30s block contains content related specifically to the scent being presented. The clips contained documentaries, cookery programs and news shows, and were chosen and altered such that the middle 30s segment corresponded to the content relating to the olfactory media [16]. Each of the six scents chosen also matched those used in the work of [16]. The scents of flowery, foul, fruity, burned, resinous and spicy reflect a "fair distribution ration between what can be termed as pleasant and unpleasant smell categories". These scents are widely used in olfactory research [13][14]. Using the same videos and scents enables a direct comparison between the work of [16] which tested olfactory and audiovisual media and this work. This comparison is discussed in detail in section 6.4.

5. Assessment Methodology

On arrival to the meeting room (room C in Fig. 3), assessors were provided with an information sheet documenting what was involved in participating in the tests. Any questions were addressed and assessors were required to sign a consent form. From here they were brought to the experimentation room (room B in Fig. 3) where testing was carried out. Assessors were told that they would see six video clips twice and that the first time they saw each clip it was "the reference sample". They were told that the second time they saw each clip it was the "sample under test". They were requested to "answer the questionnaire on their experience of the sample under test".

Table 1. Rating scales for each of the statements/questions (Likert Scale)

Score	Statement 1	Question 2	Statement 3,4,5
5	Too Late	Imperceptible	Strongly Agree
4	Late	Perceptible but not annoying	Agree
3	Neither Early or Late	Slightly annoying	Neither Agree or Disagree
2	Early	Annoying	Disagree
1	Too Early	Very annoying	Strongly Disagree

5.1 Questionnaire and Rating Scale

A number of approaches exist in the literature for offline subjective evaluations of multimedia applications. The absolute category rating (ACR) method proposed in BT.500 [39] requests participants to provide an ACR score from 1 to 5 (5 being best) after observing a single sample. With this approach, there is no reference sample and scores are given based on user expertise. This leads to non-uniform distributions of rating scores, which can invalidate subjective results [18]. Specifically in relation to olfactory media, considering the variable perception of olfactory media, this issue is exaggerated. In addition, feedback from assessors during preliminary testing indicated that the "novelty" of olfactory media made even large errors temporarily acceptable.

ITU-T P. 910 [40] proposes an alternative assessment method to address the reliance on assessor expertise by exposing participants to two media samples of different qualities and giving a comparative rating score. The first stimulus presented in each pair is always the source reference, while the second stimulus is the stimulus under test. This method is known as Degradation Category Rating (DCR) or Double Stimulus Impairment Scale method. To address the two issues highlighted above, this method was selected for the subjective testing. The reference sample was always a synchronized presentation of olfactory and video media. The samples being tested included inter-media skew of varying degrees (shown in Table. 2) as well as the synchronized presentation of olfactory and video media. For all questions, assessors chose one from the Likert scale, shown in Table 1.
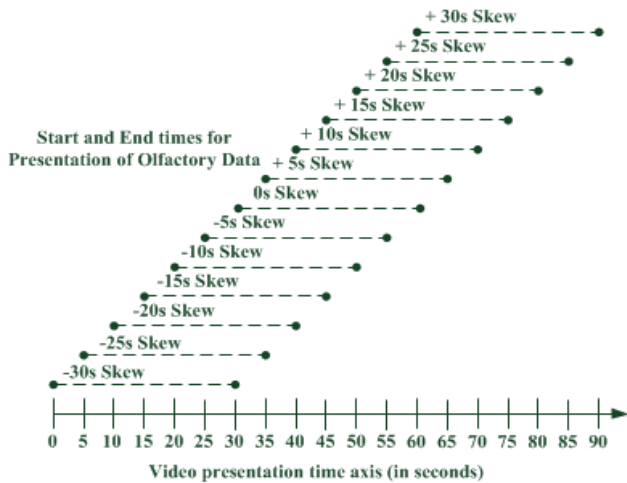
The questions used have evolved from those asked in [17] and [16]. As part of the preliminary testing, a reliability assessment was performed on the questionnaire to ascertain if the purpose and phraseology were clear and comprehendible to assessors. Discussion with each assessor was undertaken, feedback was recorded and necessary amendments were made to the draft questions. The questions explained in the remainder of this section are the final versions updated after feedback comments were considered and following review by a Psychologist. Assessors were asked to select one of the five possible answers per question as per table 1 relative to their experience of the stimulus under test.

The first statement aimed to determine assessor ability to **detect** the existence of a synchronization error, "*Relative to the content of the video clip, the smell was released:*". Assessors answered by selecting one of the five possible answers as shown under statement 1 in Table 1. Question 2 aimed to determine how **tolerable** assessors were to different levels of skew. Hence they were asked to qualify their annoyance of the inter-media skew by answering; "*In the event that you may have perceived the video clip and smell being out of synch, please indicate the extent to which it impacted upon you. Please select the appropriate option below that reflects how you would qualify it?*" As per answers for question 2 in table 1, assessors had the option of selecting one of five values that reflected how they perceived the synchronization error (if it existed) in terms of its annoyance. The mean opinion score (MOS) of respondents was used to determine the tolerable level of skew as well as deriving a level of annoyance graph as shown in Fig. 7.

The final three statements were included to analyze the impact of inter-media skew on the user experience. Assessors were asked to select one of five possible answers (see table 1, scale for statements 3, 4, 5) in terms of their agreement with the statements. The statements were ordered from general to being more specific. To determine the impact of inter-stream skew, assessors' agreement with "*You enjoyed watching the video clip*" evaluates assessor level of **enjoyment** of olfactory data as a media when in sync and explores any deterioration in this perception with the introduction of inter-media skew. "*The smell when presented, was relevant to what I was watching*" queried the **relevance** olfactory media had to the video when skews existed as opposed to

Table 2. Subject case 1 applies to participants 1, 14, 27, etc., subject case to participants 2, 15, 28 and so on

Case	Clip 1 Skew	Clip 2 Skew	Clip 3 Skew	Clip 4 Skew	Clip 5 Skew	Clip 6 Skew	Clip 1 Skew	Clip 2 Skew	Clip 3 Skew	Clip 4 Skew	Clip 5 Skew	Clip 6 Skew
1	0s	-30s	-25s	-20s	-15s	-10s	-5s	+5s	+10s	+15s	+20s	+25s
2	+30s	0s	-30s	-25s	-20s	-15s	-10s	-5s	+5s	+10s	+15s	+20s
3	+25s	+30s	0s	-30s	-25s	-20s	-15s	-10s	-5s	+5s	+10s	+15s
4	+20s	+25s	+30s	0s	-30s	-25s	-20s	-15s	-10s	-5s	+5s	+10s
5	+15s	+20s	+25s	+30s	0s	-30s	-25s	-20s	-15s	-10s	-5s	+5s
6	+10s	+15s	+20s	+25s	+30s	0s	-30s	-25s	-20s	-15s	-10s	-5s
7	+5s	+10s	+15s	+20s	+25s	+30s	0s	-30s	-25s	-20s	-15s	-10s
8	-5s	+5s	+10s	+15s	+20s	+25s	+30s	0s	-30s	-25s	-20s	-15s
9	-10s	-5s	+5s	+10s	+15s	+20s	+25s	+30s	0s	-30s	-25s	-20s
10	-15s	-10s	-5s	+5s	+10s	+15s	+20s	+25s	+30s	0s	-30s	-25s
11	-20s	-15s	-10s	-5s	+5s	+10s	+15s	+20s	+25s	+30s	0s	-30s
12	-25s	-20s	-15s	-10s	-5s	+5s	+10s	+15s	+20s	+25s	+30s	0s
13	-30	-25s	-20s	-15s	-10s	-5s	+5s	+10s	+15s	+20s	+25s	+30s

Fig. 4. Video and olfactory media presentation times during subjective testing.

Fig. 5. Detection instant per scent average and maximum/minimum detection instants per scent.

synchronized presentation. By examining the assessors' agreement with "*The smell contributed to a heightened sense of reality whilst watching the video clip*", the aim was to determine the impact the level of skew has on assessors' sense of **reality** of an olfaction enhanced multimedia clip.

## 5.2 Introduction of Artificial Skews between Olfactory and Video Media

In order to determine the perceptible and tolerable levels of inter-media skew between olfactory and video media, assessors were presented with varying levels of skew (including no skew) and queried about their perception of the experience. The audio from these video clips was removed using Windows Live Moviemaker

and replaced with the sound of a blowing fan. The sound was added to negate the influence of the noise of blowing from the SBi4 v2 (which differed depending on which fan was running). Table 2, shows the skews introduced for each of the video clips and how it was divided across participants. Once the presentation of the olfactory media was complete, a SBi4 fan with no odor cartridge was turned on to address scent lingering.

Fig. 4 shows how the olfactory media is presented at different times relative to the video clip. For olfactory media to be in sync (0s skew) with the video, it should be presented for the middle 30s block (i.e. from time 30s to time 60s on the video presentation time axis). Olfactory data before video content is represented by skew times of -30s, -25s, -20s, -15s, -10s and -5s and olfactory data after video content is represented by skews of +5s, +10s, +15s, +20s, +25s and +30s.
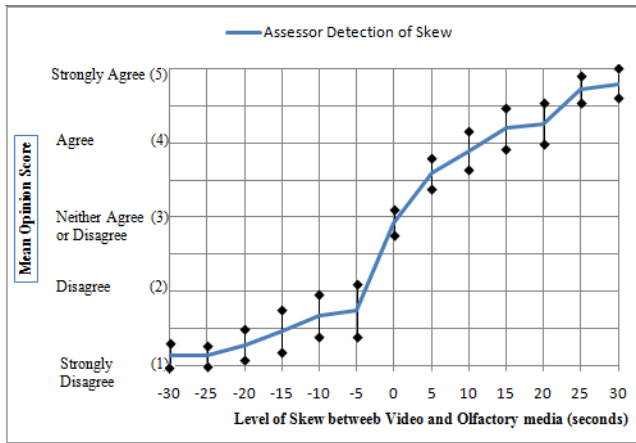
Fig. 6. Analysis of skew detection with confidence interval based on 99% confidence level.

Assembling results to these questions under the conditions discussed provides a holistic view of the detection and perception of inter-media skew of multimedia applications based on olfactory data and video. It also enabled us to gather data on the impact of inter-stream skew on the user QoE. In the next section of this paper, the results from the experiments are presented, explained and discussed.

6. Results and Discussions

The results of the subjective testing discussed above are presented and the definition of temporal boundaries for olfactory and video synchronization is explained.

6.1 Preliminary Experiment: Measurement of the Detection Instant.

Because of the slow moving nature of olfactory data compared with audio or video media, it was critical for the synchronization study to determine how long it took assessors to detect the presence of odors once emitted. 15 participants (9 male, 6 female) were presented with the 6 scents twice in random order. Assessors clicked on the mouse once they detected a scent. As we considered it took 1 second for assessors' reaction and click on the mouse we determined, on average, but per scent, how long in advance the olfaction device's fans should be started in order to ensure timely presentation to the users. Based on the SBi4 being 0.5 meters from the assessor, it was found that it took assessors between 2.7s - 3.7s to detect the scents depending on the scent as per Fig. 5.
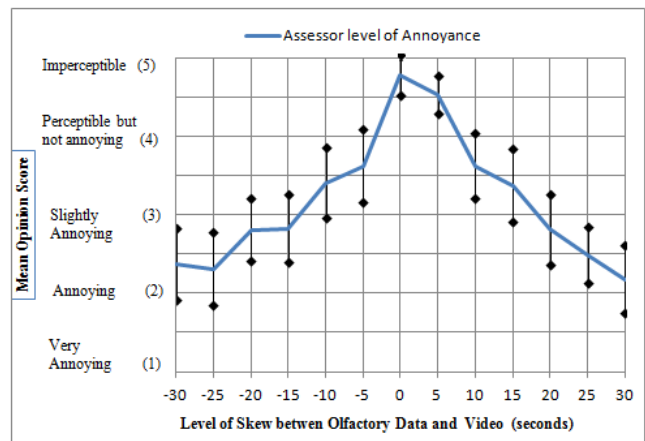
6.2 Detection and Perception of Error

Figure 6 gives an overview of the results of statement 1, to determine users' ability to **detect** levels of inter-media skew. The vertical axis shows the five possible answers to question one i.e. when the scent arrived relative to the video. The horizontal axis indicates the level of skew artificially introduced between the olfactory and video media with the negative values representing olfactory media before video media. Analysis of Fig. 6 shows that assessors were able to identify the existence of inter-stream skew very well. It also indicates that assessors were much more sensitive to scent that was early rather than late based on the comparison of skews before and after time 0s. Direct comparison of MOS scores at skews of +5s and -5s show that the MOS for +5s of 3.59 was much closer to being at the "correct time"

(represented by a value of 3) as opposed to the value of 1.74 for -5s. Interestingly based on MOS comparison, assessors viewed skews of +15s and -5s similarly in terms of being Late or Early respectively. In order to analyze if significant differences existed in participants' perception between synchronized and unsynchronized scent and video, the data collected was analyzed using paired sample t-test. With 99% confidence level, the t-tests showed for all levels of skew between the olfactory data and video that the significant two tailed p values were less than 0.01 (p<0.01) (+5s had a two tailed p value of 0.0000001, -5s had a two tailed p value of 0.000000007) and hence it can be concluded that there is a statistically significant difference between the mean of the synchronized and mean of participant responses for the "skewed" release times for detection of skew.

The task of question 2 was to determine the effect that the skews had on the perception of the olfactory-video clip. The effect an error has is key to determine temporal boundaries, as works involving other media have shown that users can tolerate certain levels of skew [16] [17]. Hence, assessors were asked to qualify the level of impairment the inter-media skew had on the experience when comparing it to the synchronized reference sample. Fig. 7 shows the MOS for level of annoyance for inter-media skew. Scores of 1 or 2 or 3 represent the perception of the skew being annoying with scores of 4 and 5 representing a tolerable skew level and imperceptible error respectively. In the next section we discuss how these results are interpreted to define temporal synchronization boundaries of olfactory data and video.

As was the case with data collected for statement one, a paired sample t-test was run with a confidence interval of 99% assessing significant differences in assessors' perceptions during the perfectly synchronized case and the cases where skews existed. For skew size of +5s, the results between an in-sync and skewed presentation were found not to be statistically significant, with the significant two tailed value of 0.160 which is greater than 0.01 (p>0.01). For all other levels of skew between the olfactory data and video that the significant two tailed values were less than 0.01 (p<0.01) (-5s had a two tailed p value of .00000001, +10s had a two tailed p value of .000002), hence it can be concluded that there is a high statistically significant difference between the mean of the case of synchronized presentation in comparison with the "skewed" presentations except skew of +5s.

Fig. 7. Analysis of annoyance level per skew with confidence interval based on 99% confidence level.

6.2.1 Detection and Perception of Synchronization Error Discussion

Findings from Fig. 6 support the plethora of literature that recounts human sensitivity to, and ability to, detect scent. If we consider and compare the MOS values for corresponding skews i.e. -30s with +30s, -25s with +25s, -20s with +20s, -15s with +15s, -10s with +10s and -5s with +5s, interesting conclusions can be drawn. For skew sizes of greater than +25s or -25s, the differences in the assessor's detection are generally similar. However, for skews of video -20s and +20s, the MOS scores indicate greater sensitivity to olfaction before video, than video before olfaction. This is reflected in the very close scores between -30s, -25s and -20s (almost same MOS rating i.e. both detected as almost "very early" 1.13s, 1.13s 1.28s respectively). If the +20s value is analyzed, its rating is not as close to the "too late" rating as the -20s rating is to "too early". In fact the MOS rating (4.26) is much closer to "late" than "too late". This is exaggerated at the corresponding 15s skews (i.e. -15s and +15s and 10s skews (i.e. -10s and +10s). Olfaction before video with a skew of -15s has a MOS score of 1.46 or hence being classified as being closer to "too early" than "early". The rating for olfaction after video by the same amount has a value of 4.2 (much closer to late which has a score of 4 than "too late" which would be reflected by a score of 5). Olfaction before video with a skew of -10 s has a MOS score of 1.67 or hence being classified as being between "too early" and "early". The rating for olfaction after video by the same amount (i.e. +10s) has a rating of 3.9 (as opposed to a MOS score of 4 which would be expected for a skew that is "late"). As is clear from the Fig. 6, this trend continues for skews of +5s and -5s. Considering this, it is plausible to conclude that olfaction **before** video is more **noticeable** than olfaction **after** video.

In terms of rating the impairment caused by the existence of a synchronization error, Fig. 7 details the assessor rating of the skews according to the degradation category rating scale. At large skews of -30s and + 30s, the MOS scores reflect equal measures of annoyance (both close to the "annoying" rating) with MOS of 2.38 and 2.18 for -30s and +30s respectively. Comparing MOS scores for skews of +25s/-25s show similar ratings with early slightly more annoying with in terms of annoyance with a score 2.31 as opposed to 2.49. Analyzing scores for skews of -20s and +20 s, again shows little or no difference with both being approximately 2.8, reflecting this impairment as between "annoying" and "slightly annoying". The first significant change in scores between scent before or after the video occurs at skews of 15s. With scent 15s before video (or -15s) a rating of 2.83 exists. The score is very similar to the rating for -20s and reflects this experience as being between annoying and slightly annoying. Comparing this with the equivalent skew when olfaction is after the video, the MOS score of 3.38 is in the range between "slightly annoying" and "perceptible but not annoying". This is highlighted further when analysis of the -10s and +10s skews is performed. The MOS rating for -10s skew is 3.41 which is in between the "perceptible but not annoying" to "slightly annoying" range. The MOS value for +10s skew reflects that this skew is also in the "perceptible but not annoying" to "slightly annoying" range although with a score of 3.63 with scent after video marginally less annoying that scent before video. Another interesting result occurs at skews of 5s. Scent presented -5s before video has the same scoring rating as when scent is presented +10s after video, with a score of 3.63, still in the "perceptible but not annoying" to "slightly annoying" range. However, scent presented +5s is in the "perceptible but not annoying" to "imperceptible" range with a

score of 4.54. Hence, it is valid to conclude, based on the comparison of MOS scores with olfaction before and after video, that assessors are **less tolerable** to olfaction **before** of video than they are of olfaction **after** video.

In terms of defining the temporal boundaries for synchronizing olfactory and video media, these results reveal "in-sync" and "out-of-sync" regions. These boundaries are based on (1) the above results that indicate assessors are more tolerable to skews when olfaction is after video and (2) A impairment rating of above 3.5 i.e. between "perceptible but not annoying" and "slightly annoying" is the minimum standard to be adhered to in order to support synchronized presentation of olfactory and video media, (3) do not consider differences in age or gender, then:

- The in-synch region spans between a maximum skew of -7.5 seconds when olfaction is before video, and a maximum skew of +10 s when olfaction is after video.

- The out-of-synch region for olfaction before video spans beyond the skew of -7.5 s and skew of greater than +10s when olfaction is after video

The in-synch region is based on the range where assessors perceive errors to be not annoying (rating above 3.5) as per Fig. 7. Considering the assessor tolerance to olfaction **after** video as opposed to olfaction before video, the span is larger for olfaction after video. This is plausible, in our everyday lives; we see first, smell later.

6.3 Impact of Synchronization Error on User Experience

This section analyzes the impact inter-media skews have on the QoE of olfaction enhanced video clip. To address this, assessors were asked to choose one of five levels of agreement with statements 3, 4 and 5 already outlined in section 5.1. The MOS results presented in the following sections compare assessors' experience of olfactory and video media with the existence of skew as against the case of synchronized presentation in terms of: sense of relevance, sense of reality and enjoyment.

6.3.1 Impact on Sense of Relevance

To determine the impact of skew on the perceived sense of relevance, assessors were required to express their level of agreement with statement 3 as outlined in section 5.1. Fig. 8 shows the MOS reflecting assessors' level of agreement with statement 3 in the presence of varying degrees of inter-media skew. When synchronized presentation takes place, assessors agreed that the smell was relevant to what they were watching. Also, it is clear that in the presence of large skews, e.g. -30 s or +30 s the sense of relevance is affected as the MOS values show that generally participants neither agreed nor disagreed with statement 3. Interestingly, the level of agreement deteriorated much more quickly as the levels of skew increased with olfaction **before** video as opposed olfaction **after** video. Fig. 8 shows a sharp increase in sense of relevance between skews of -15s to -10s with MOS increasing from 2.5 to 3.56. Olfaction **after** the video is more relevant than olfaction **before** video. Olfaction with a skew of +15s (15s after video) is still more relevant than scent with a skew of -5s. The slow gradual decrease in relevance for olfaction **after** video is particularly interesting when it is compared with the sharp decrease the larger the skew size when scent is **before** video.

A paired sample t-test with a 99% confidence interval was used to determine if statistically significant differences existed in responses between synchronized and unsynchronized presentations of scent and video. For skew size of +5s, +10s, +15s and +20s the results were not significantly different with the significant two tailed values of 0.148, 0.146, 0.036 and 0.037 respectively ($p>0.01$). For skew size of -5s and -10s the results were not significantly different with the significant two tailed values of 0.74 and 0.47 respectively, again greater than 0.01. For all other skews the significant two tailed values were less than 0.01 ($p<0.01$), (-15s had a two tailed p value of 0.0000007, +25s had a two tailed p value of 0.001, less than 0.01 ($p<0.01$)) hence all other differences were statistically significant.
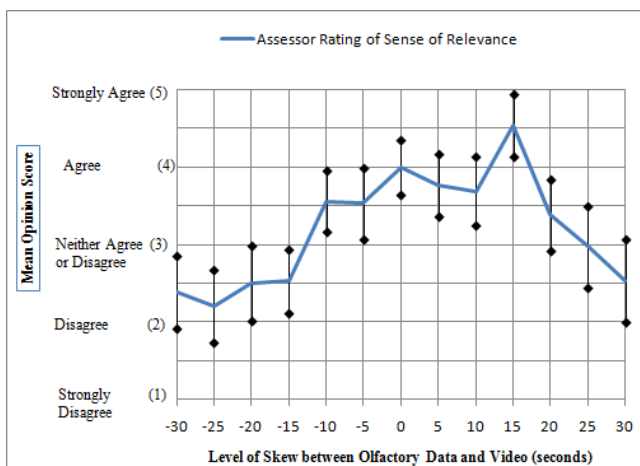

Fig. 8. Analysis of relevance per skew with confidence interval based on 99% confidence level.

6.3.2   Impact on Sense of Reality

Fig. 9 shows MOS scores reflecting assessor's level of agreement with statement 4 in the presence of varying degrees of inter-media skew. When synchronized presentation takes place, assessors agreed that the smell heightened the sense of reality of what they were watching. However, again in the presence of large skews (e.g. -30 s or +30 s), the MOS values show that generally participants were between "Neither agree or Disagree" and "Disagree" with statement 4 in the presence of skew. For olfaction before video, with skews of -10s and -5s, assessors perceived the olfaction somewhat contributing to an enhanced sense of reality with a MOS of 3.46 between "Neither agree or Disagree" and "Agree". With no skew, 0s, and +5s skew assessors "Agree" that the scent contributed to a heightened sense of reality. The most interesting finding from analysis of the MOS from this question was the **slow reduction** in heightened sense of reality for olfaction **after** scent with skews of +5s, +10s and +15s. Just 0.15 separates the rating of +5s from +15s. When comparing the ratings of scent before or after the video, the opinions in terms of impact on reality were that not much difference existed between scent -5s before video and +20s.
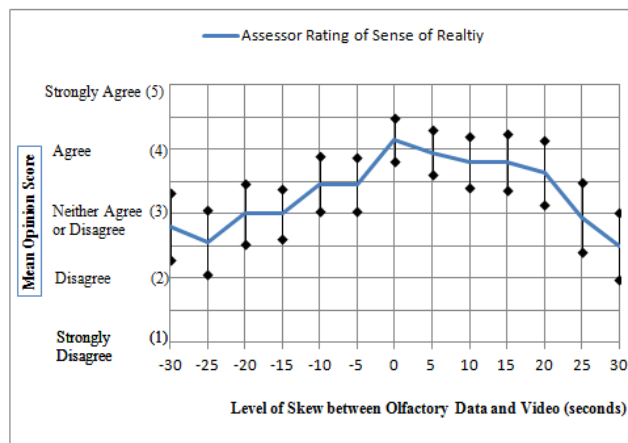

Fig. 9. Analysis of Sense of reality per Skew with confidence interval based on 99% confidence level.

Assessors' opinions were compared to determine if statistically significant differences existed in responses when a synchronization error was present with the case when olfaction and video were perfectly synched via a paired sample t test. For skew size of +5s, +10s and +15s the results were not significantly different with the significant two tailed values were 0.163, 0.22 and 0.013 respectively which are greater than 0.01 ($p>0.01$). For all other skews the significant two tailed values were less than 0.01 ($p<0.01$) (+20s had a significant two tailed value of 0.002, -5s had a significant two tailed value of 0.005), hence the results for skewed presentation were statistically significant when compared with ratings for in-sync presentation.

6.3.3   Impact on Sense of Enjoyment

Figure 10 shows for MOS reflecting assessor's level of agreement with statement 5 in the presence of varying degrees of inter-media skew. When synchronized presentation takes place, assessors agreed that they enjoyed watching the video clip. In the presence of large skews (e.g. -30 s or +30 s), the MOS values show that generally participants were between "Neither agree or Disagree" and "Disagree" with statement 5 in the presence of large skew. Between skews of -10s and +10s participants were between "Neither agree or Disagree" and "Agree" with the highest levels of agreement existing at 0s and +5s skew with scores of 3.98 and 3.84 respectively. At skew level of -5s, assessors also rated the enjoyment highly with a score of 3.58. Outside skews of -10s and +15s participants scores reflected that they "Neither Agree or Disagree" with the sense of enjoyment statement.

Assessors' opinions were compared to determine if statistically significant differences of means existed in responses when a synchronization error was present with the case when olfaction and video were perfectly synched via a paired sample t test with a confidence interval of 99%. For skew size of +5s and -5s the results were not statistically significant from the synchronized case with the significant two tailed values were 0.361 and 0.49 respectively which is greater than 0.01 ($p>0.01$). For all other skews the significant two tailed values were less than 0.01 ($p<0.01$), hence the mean results were statistically significant for all skews except -5s and +5s.
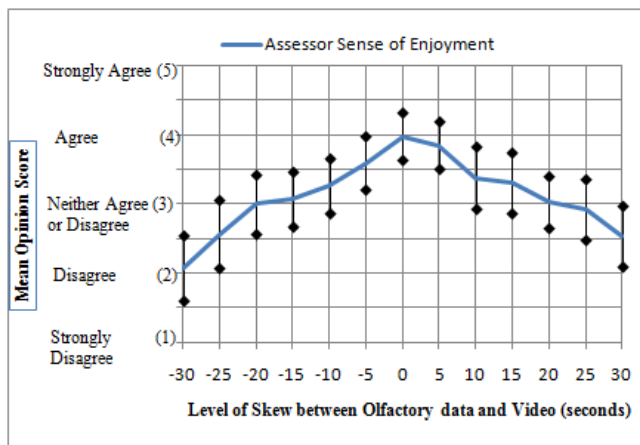
Fig. 10. Analysis of sense of enjoyment per skew with confidence interval based on 99% confidence level.

6.4 Comparison of Audiovisual & Olfactory with Video and Olfactory User Perceived Synchronization

This section compares this work with current state of art research on olfactory and audiovisual media synchronization, highlighting the similarities and differences between them. As already mentioned, the same videos were used in both works. In [16][29] and here the middle block of approximately 30s related to the scent being presented. The videos were edited such that the audio was replaced with the sound of a blowing fan and not contextual audio data associated with the meaning of the clip. Also the same smells were used although, they were provided by different vendors. [16] states that it took "about 2 seconds" for emitted scent to reach users using the Vortex Active emitter. As discussed above, it was experimentally evaluated that it took between 2.7s – 3.7s using the SBi4 emitter. Artificial skew values ranging from -30s to +30s were tested in [16] in step sizes of 10s (hence each assessor tested 6 samples), whereas the step sizes here were of magnitude of 5s (each assessor tested 12 samples). In both works the playback order was randomized to minimize ordering effects. Approximately the same number of participants took part in each experiment (42 in [16] as opposed to 43 here).

In terms of conclusions, [16] define that the temporal relationship between olfactory and audiovisual media is from -30s to +20s. With the removal of contextual audio, our results show an "in-sync" region of –7.5s to +10s. The difference is surprisingly large, but as found during our initial tests; the presence of related audio validates the findings of [16]. Assessors noted that the scent, when complemented by relevant and contextual audio, provided extra information even with large skews compared to the video content. Without the presence of contextual audio, assessors, as shown by this work, assessors found it confusing, annoying and it had negative impacts on relevance, sense of reality and enjoyment. Finally, [16] found that olfaction before audiovisual information was more tolerable than olfaction after audiovisual media. As shown above, the **reverse** is true for olfaction and video only i.e. olfaction after video **only** is more tolerable.

7. Conclusion and Future Work

This work investigated the perception of inter-stream synchronization error between olfactory and video media and its affect on the quality of experience for the user. It defined user temporal boundaries for user perceived synchronization. Synchronization is supported up to a maximum skew of 7.5 seconds when olfaction is before video, and a maximum skew of 10 seconds when olfaction is after video. Skews outside this range are in the out-of-sync region. Interestingly, this region is significantly smaller than the equivalent boundary for synchronization of olfaction, video and audio. That temporal boundary spans 30 seconds when olfaction is before audiovisual media and a maximum of 20 seconds when olfaction is after audiovisual media. In addition, the results indicate that assessors were more aware of olfaction before video as opposed to olfaction after video. It was also identified that assessors are less tolerable to olfaction before video than they are of olfaction after video. Finally, this work also analyzed the impact of inter-stream skew on assessors QoE. It was found that skews beyond a particular range between olfaction and video impact the sense of relevance, reality and enjoyment negatively when compared with feelings achieved when synchronized presentation took place.

There are many possibilities for future work related to olfactory data as a media. The next steps involve analysis of how age and gender affect perceived QoE. On completion of this, we aim to deduce an analytical model that results in user profiles.

9. REFERENCES

[1] Cha, J., Eid, M., Barghout, A., Mahfujur Rahman, ASM., and El Saddik, A. 2009. HugMe: Synchronous Haptic Teleconferencing. In *Proceedings of 17th ACM International Conference on Multimedia* (ACM MM '09).

[2] Narumi, T., Nishizaka, S., Kajinami, T., Tanikawa, T., and Hirose, M. 2011. Augmented Reality Flavors: Gustatory Display Based on Edible Marker and Cross-Modal Interaction. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (CHI '11).

[3] Ghinea, G. and Ademoye, O. A. 2012. The Sweet Smell of Success: Enhancing Multimedia Applications with Olfaction. In *ACM Transactions on Multimedia Computing, Communications and Applications* 8, 1 (Jan. 2012).

[4] Hoshino, S., Ishibashi, Y., Fukushima, N., and Sugawara, S. QoE Assessment in Olfactory and Haptic Media Transmission: Influence of Inter-Stream Synchronization Error, 2011. In *IEEE Int. Workshop Technical Committee on Communications Quality and Reliability* (CQR '11).

[5] Timmerer, C., Waltl, M., Rainer, B., Hellwagner, H. 2012. Assessing the quality of sensory experience for multimedia presentations. In *Signal Processing: Image Communication* 27, 8, pp 909-916 (Feb. 2012).

[6] Tomono, A., Yamamoto, S., Utsunomiya, M., Ikei, D., Yanagida, Y., Hosaka, K.. 2004. Effect that the Image Media with scent Gives to Contents Understanding. In *Human Interface Symposium* pp. 249-254 (March 2004).

[7] Nakamoto, T., Otaguro, S., Kinoshita, M., Nagahama, M., Ohinishi K., Ishida T., 2008. Cooking up an Interactive Olfactory Game Display. In *IEEE Computer Graphics Applications* 28, 1 pp 75-78. (Jan-Feb 2008).

[8] Spencer, B. S., "Incorporating the Sense of Smell Into Patient and Haptic Surgical Simulators" In *IEEE TRANSACTIONS on Information Technology in Biomedicine* 10, 1, 168-173 (Jan. 2006).

[9] Shams, L., Seitz, AR, (2008). Benefits of Multisensory learning. In *Trends in Cognitive Sciences* 12,11 pp 411-417 (Nov. 2008)

[10] Washburn, D.A., "Olfactory Use in Virtual Environment Training". In *Modelling and Simulation* 2, 3, pp 19-25 (2003).

[11] Dann, G., Jacobsen J., K., S., "Tourism smellscapes" In *Tourism Geographies*, 5, 1, pp 3-25 (Feb. 2003)

[12] Wysocki, C.J., Pierce, & Gilbert. "The Future of Television: Sweeping Change at Breakneck Speed 10 Reasons You Won't Recognize Your Television in the Not-Too-Distant Future, " Cisco Internet Business Solutions Group 2011.

[13] Chastrette, M., 2002. Classification of Odors and Structure – Odor Relationships. In *Olfaction, Taste, and Cognition* pp. 100-116, Cambridge University Press (2002).

[14] Kaye, N., "Symbolic olfactory display" M.S. thesis, Massachusetts Inst. Techonology, Cambridge, MA, 2001.

[15] Huang, Z., Wu, W., Nahrstedt, K., Rivas, R., Arefin, A., 2011. SyncCast: Synchronized Dissemination in Multi-site Interactive 3D tele-immersion, In *Proc of 2nd annual ACM conference on Multimedia Systems*, pp 69-80 (2011).

[16] Ghinea, G. and Ademoye, O. A., 2010. Perceived Synchronization of Olfactory Multimedia. In *IEEE Trans. on SYSTEMS, MAN, AND CYBERNETICS – PART A: SYSTEMS AND HUMANS* 40, 4, 657-663 (July. 2010).

[17] Steinmetz, R. 1996. Human Perception of Jitter and Media Synchronization. In *IEEE Journal on Selected Areas in Communications* 14, 1 61-72 (Jan. 1996).

[18] Huang, Z., Nahrstedt, K., 2012. Perception-based playout scheduling for high-quality real-time interactive multimedia. In *Proceedings of IEEE International Conf. on Computer Communications (INFOCOM)* pp. 2786-2790 (Mar. 2012)

[19] Huang, Z., Arefin, A., Agarwal, P., Nahrstedt, K and Wu, W., 2012. Towards the understanding of human perceptual quality in tele-immersive shared activity. In *Proc. of ACM Multimedia Systems Conf. (MMSYS),* pp 29-34 (Feb. 2012).

[20] Gulliver, S. R., Ghinea, G., 2007. Defining the users perception of distributed multimedia quality. In *ACM Trans. on Multimedia Computing, Communications and Applications (TOMCCAP)* 2, 4, pp 241-257 (Nov. 2007).

[21] Lee, J.-S., Goldmann, L., and Ebrahimi, T., 2011. A new analysis method for paired comparison and its application to 3D quality assessment. In *Proc. of ACM International Conference on Multimedia* pp1281-1284 (MM'11).

[22] Ghinea, G., Ademoye, O.A., 2011. Olfaction-enhanced multimedia: perspectives and challenges. In *Multimedia Tools Appl* 55, 3, pp. 601-626 (Dec. 2011)

[23] Ramic, B., Hasic, J., Rizvic, S., Chalmers, A. 2006. Selective Rendering in a multimodal environment: Scent and graphics," In *Spring Conference on Computer Graphics, ACM SIGGraph,* (2006).

[24] Nakamoto, T., Yoshikawa, K., 2006. Movie with Scents Generated by Olfactory Display Using Solenoid Valves. In *Proceedings of Virtual Reality Conference* pp 291-292 (March 2006).

[25] Arefin, A., Huang, Z., Rivas, R., Shi, S., Wu, W., Nahrstedt, K. 2011. Tele-immersive Gaming for Everybody. In, *Proc of 19th ACM Int. Conference on Multimedia* pp 783-784 (2011).

[26] Zhou, J., Malric, F., Petriu, E.M., Georganas, N.D. 2011. Uniform Hardness Perception in 6-DOF Haptic Rendering. In *IEEE Trans. Instrum. Meas.* 60, No. 1, pp. 214-225 (Jan. 2011).

[27] Eid, M., Cha, J., and Saddik, A. El,,, "Admux: An Adaptive Multiplexer for Haptic-Audio-Visual-Data Communication" In *IEEE Trans. Instrum. Meas*, 60 1, 21-31 (Jan. 2011).

[28] ISO/IEC FDIS 23005-3 Information Technology – Media context and control- Part 3: Sensory Information, ISO Publication (2010)

[29] Ademoye, O. A., Ghinea, G., 2009. Synchronization of Olfaction-Enhanced Multimedia. In *IEEE Transactions on Multimedia* 11, 3, pp 561-565 (April 2009).

[30] Steinmetz, R., Nahrstedt, K., 1995. Multimedia: computing, communications and applications. In *Prentice Hall.* (Sept. 1995).

[31] Exhalia www.exhalia.com (accessed 03.01.2013)

[32] ISO/IEC 8589 Sensory analysis – General guidance for the design of test rooms.

[33] Nakamoto, T., Minh, H.P.D., 2007. Improvement of Olfactory Display Using Solenoid Valves. In *IEEE Virtual Reality* pp 179-186, (March 2007).

[34] Ariyakul, Y., Nakamoto, T., 2011. Olfactory display using a miniaturized pump and a SAW atomizer for presenting low-volatile scents. In *IEEE Virtual Reality* pp 193-194, (March 2011).

[35] Sugimoto, S., Noguchi, D., Bannai, Y., Okada, K., 2010. Ink jet olfactory display enabling instantaneous switches of scents. In *Proceedings of ACM International Conference on Multimedia* pp 301-310 (October 2010).

[36] Noguchi, D, Sugimoto, S., Bannai, Y., Okada, K., 2011. Time characteristics of olfaction in a single breath. In *Proceedings of the ACM SIG CHI Conference on Human Factors in Computing Systems* pp 83-92 (CHI '11).

[37] ISO 5492:2008 Sensory analysis – Vocabulary, International Standards Organization (ISO) publication 2008 as technical revision of ISO 5492:1992

[38] Lawless, H. T. 1997. Olfactory Psychophysics. In *Tasting and Smelling. Handbook of Perception and Cognition, 2nd Edition* pp125-174. Beauchamp & Bartoshuk (Eds), Academic Press, San Diego, CA (1997).

[39] ITU-T BT.500. Methodology for the subjective assessment of the quality of television pictures, 2002.

[40] ITU-T P.910. Subjective video quality assessment methods for multimedia applications, 2008.

[41] Calvert, G., Spence, C., Stein, B.E. *The Handbook of Multisensory Processes*. In MIT Press. 2004