Institution: The University of Essex		
Unit of Assessment: 11 – Computer Science and Informatics		
Title of case study: New techniques for measuring speech and video to enhance the quality of Internet streamed media		
Period when the underpinning research was undertaken: 01.01.2000 – 6.10.2009		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Prof Martin Reed	Professor	1998 – Present
Prof Mohammed Ghanbari	Professor (Emeritus)	1986 – July 2011 (Emeritus since August 2011)
Prof Rob Massara	Professor (Emeritus)	1978 – July 2011 (Emeritus since August 2011)
Period when the claimed impact occurred: 01.08.2013 – 31.07.2019		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact

Essex researchers have pioneered new techniques for measuring the quality of speech and video transmitted over the Internet. A key contribution to the quality measurement was a new time-alignment method developed at Essex, in collaboration with Psytechnics. This work was incorporated into Microsoft's product Skype for Business and by 2016 was used to measure the voice quality of approximately 100 Million users globally, including the US and UK, thus, improving the quality of teleconferencing. The same time-alignment technique was critical in the video quality measurement standard (J.247 PEVQ). PEVQ has been used worldwide as the standard manner that Internet-based video systems have been measured since 08/2008 to improve video delivery services, and is still in use today. A new non-intrusive speech measurement system, developed by Essex in collaboration with Psytechnics, led directly to the standard P.563. This work formed a key product for Netscout and contributed USD500 Million annual revenue to the company by 2018. Through this product, Essex research was used to monitor networks of service providers, government agencies, large financial institutions and other enterprises across the globe to improve the quality of speech communication.

2. Underpinning research

Research from the University of Essex has created new measurements of the quality of speech and video transmitted over the Internet. This work was needed because of the transition of mass media transmission from traditional broadcast TV and telephone systems to the new packet networks (e.g., the Internet), which started around the year 2000, but which is now ubiquitous. This change required completely new methods to measure the quality of speech and video transmission that were pioneered by University of Essex research. In particular, two such developments are described below, the first considered the time alignment of speech and video so that they can be

measured accurately in packet-based (Internet) transmission that is inherently asynchronous **[R1, R2, R3]**. The second innovation considers how a speech transmission system (such as the Internet) can be measured without disturbing the conversation in progress; this is termed a "non-intrusive" measurement **[R4]**. All of this research was carried out collaboratively with Psytechnics **[S2, S7]** and appears in peer-reviewed international conferences or journals with more than 150 citations.

Time alignment methods for speech and video quality measuring systems:

A strategic piece of work by the University of Essex, carried out by Reed **[R1]** and Ghanbari **[R2, R3]**, in collaboration with Psytechnics **[S2]**, investigated the problem of time-alignment in measuring media quality transmitted over the Internet. In Internet streaming, data is split into individual packets that may be subject to arbitrary delay. This was a problem for existing measurement systems that relied upon the received signal being exactly aligned with the original. University of Essex researchers addressed the issue by pioneering practical statistical methods to align the media allowing the measurement of speech and video quality in the new packet-based transmission medium of the Internet. The technique relies upon creating a histogram of audio or video temporal events and using this to compare over previous/later frames to find the optimal alignment point. This work was used by Psytechnics as an essential component of their measurement products **[S2]** and for the Psytechnics contribution to the ITU-T standard for video quality measurement J.247 **[S4]** (also called PEVQ); this continues to be a standard method for measuring the quality of encoded video transmitted over packet media such as the Internet.

Non-intrusive speech quality measurement:

The non-intrusive measurement of speech quality carried out by Rob Massara **[R4]**, again in collaboration with Psytechnics **[S2, S7]**, led directly to the ITU-T standard P.563 **[S6]**. The paper **[R4]** describes an overall method for performing this new measurement which is described in Figure 8 of the standard **[S6]**, for example Equation 3 in the paper, which models the human vocal tract, is used directly in the standard **[S6, Section 9.2.2]**. This work is important, as it was the basis of the first practical implementation of a method to measure speech quality without having access to the original signal, i.e., a no-reference measurement. Previous techniques either, required adding disturbing signals during the communication or measuring the communication system (e.g., mobile phone and intervening systems) before or after the actual conversation; this meant it was not possible to measure the quality of any particular conversation while the conversation was in progress. This novel non-intrusive measurement technique models the human vocal tract and compares the measured speech with this model to discriminate between real speech and errors in the transmission. Consequently, the speech quality can be determined by simply monitoring the transmitted voice data, while the conversation is in progress, without disturbing the conversation.

3. References to the research [can be supplied by HEI on request]

[R1] L. Malfait, P. Gray and M. J. Reed, "Objective listening quality assessment of speech communication systems introducing continuously varying delay (time-warping): A time alignment issue," *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, 2008, pp. 4213-4216. doi: 10.1109/ICASSP.2008.4518584

[R2] Q. Huynh-Thu and M. Ghanbari, "No-reference temporal quality metric for video impaired by

frame freezing artefacts," *2009 16th IEEE International Conference on Image Processing (ICIP),* Cairo, 2009, pp. 2221-2224. doi: 10.1109/ICIP.2009.5413894

[R3] Q. Huynh-Thu and M. Ghanbari, "Temporal Aspect of Perceived Quality in Mobile Video Broadcasting," in *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 641-651, Sept. 2008. doi: 10.1109/TBC.2008.2001246

[R4] P. Gray, M. P. Hollier and R. E. Massara, "Non-intrusive speech-quality assessment using vocal-tract models," in *IEE Proceedings - Vision, Image and Signal Processing*, vol. 147, no. 6, Dec. 2000, pp. 493-501. doi: 10.1049/ip-vis:20000539

4. Details of the impact

The work at the University of Essex on the measurement of speech and video quality has had global impact on speech and video streaming services; these types of services contributed to 75% of the Internet traffic by 2017 and growing (estimated as 82% by 2022) **[S1]**. The first specific impact is the work by Reed **[R1]** which led to reliability measurement systems incorporated from 2006 until 2016 into the **global product Microsoft Skype for Business [S2] with more than 100 Million users [S3].** The work by Ghanbari **[R2**,**R3]** built on the work of Reed **[R1]** to provide time-alignment for video and was included in ITU-T standard J.247, also called PEVQ **[S4]**; J.247 has been **the standard manner that Internet based video systems have been measured since 08/2008 [S5, S4]**. The work by Massara **[R4]** let to the ITU-Standard P.563 **[S6],** which formed an important product for Psytechnics and later Netscout **[S7]** where it has been used by **global industry to improve the quality of speech communication [S7]**.

The time alignment work by Reed from University of Essex **[R1],** in collaboration with Psytechnics **[S2],** has become an essential part of the systems used by industry to measure and improve Internet speech communication quality and reliability. The findings of this research were incorporated into Microsoft's Skype For Business product (also called Lync) from 2006 to 2016 **[S2]**. In 2015 Microsoft stated that there were 100 Million people using Microsoft Skype for Business for their work **[S3]**, and that as '*part of the Office 365 platform, Skype for Business will deliver at **massive worldwide scale**, with datacenters in 37 countries; with the most comprehensive productivity experience; and enterprise-grade **reliability** and controls on top of a secure and compliance-ready platform'* **[S3]**. The implementors of the measurement system, formally from Psytechnics, state that the *'research from the University of Essex in the area of time alignment for the measurement of quality of service of audio and video media was an **essential component in the systems developed by Psytechnics**. Specifically, their time alignment method was used in NEXTGENPESQ, PsyVoIP products. NEXTGENPESQ and PsyVoIP were sold under license in 2006 to Microsoft, and PsyVoIP was thus integrated into Lync (later also called Skype for Business) by Microsoft for live monitoring of this product. This was an essential component for monitoring and improving the quality of Microsoft's Lync (Skype for Business) teleconferencing product and the time alignment method **originating from the University of Essex's research was used in that product until 2016**. Without such a component, Microsoft would not have been able to determine the speech quality of their product. The extent to which Essex's research on improving the quality of teleconferencing was used proved to be **truly global as Lync was included in all Enterprise editions of Microsoft Office'* **[S2]**. In 2016, an independent report found that **33% of US enterprises** (companies with more than 500 employees) were using Lync/Skype for Business **[S8, p9]**.

The time alignment work at the University of Essex by Reed **[R1]** was further developed by Ghanbari **[R2, R3]**, in collaboration with Psytechnics **[S2]**, for use with video quality measurement and included in the J.247 standard **[S2, S4]**; Psytechnics' implementation is described in Annex C of the standard (**[S4]**, Psytechnics full reference method, p.51 – 79). This objective measurement system, using the time alignment process produced by the University of Essex, replaces time consuming subjective measurements (using large panels of users), thus significantly reducing costs to industry. Opticom, who are the provider of the license of the J.247 measurement system, list 29 large multi-national companies (including Microsoft and Ericsson) from 10 different countries (including US, China, South Korea, Germany) licensed to use PEVQ (i.e. J.247) to test video quality for their products and thus improve their video delivery services **[S5]**.

The research work by the University of Essex into non-intrusive speech quality by Massara **[R4]**, '*was an essential component of Psytechnics Experience Manager and Psytechnics no-reference model was integrated into the ITU-T standard P.563'* **[S6] [S2]**. This product line was sold to Netscout in 2011 **[S7]** and was subsequently formed (after rebranding) into the Netscout product nGeniusOne in 2015 **[S7]**, which measures the quality of Internet/packet-based speech communication **[S9]**. In the form of the Netscout nGeniusOne product, the '*non-intrusive speech monitoring work, from the Essex collaboration, became an essential differentiator in Netscout nGeniusONE and was one of the reasons customers chose to invest in the solution which **had an annual revenue of approximately USD 350 Million by the end of 2016**. Netscout's nGeniusONE established itself as an important product for Netscout and **by 2018 was responsible for USD 500 Million of annual revenue for Netscout**, with the **work originating from the research undertaken in collaboration with the University of Essex providing an essential component for the speech monitoring** in Netscout nGeniusONE'* **[S7]**. This innovation from the University of Essex '*continues to be used by **mobile and fixed service providers, government agencies, large financial institutions, healthcare providers and a range of other enterprise customers from varied industries across the globe**'* **[S7]**.

5. Sources to corroborate the impact

[S1] Cisco, "Cisco Visual Networking Index Complete Forecast Highlights," 2018, https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_2022_Forecast_Highlights.pdf (Accessed 29/1/2021)

[S2] Corroboration from VP/CTO Communications Business Group, Dolby (formally CEO/CTO Psytechnics) that research from the University of Essex was included in the standards J.247 and P.563 and confirmation of the commercial impact of the time-alignment technique in the Microsoft audio conferencing tools Lync/Skype for Business.

[S3] Post from the Skype for Business Team March 18th 2015 https://www.microsoft.com/en-us/microsoft-365/blog/2015/03/18/skype-for-business-is-here-and-this-is-only-the-beginning/ (Accessed 15/1/2021)

[S4] ITU-T Standard J.247 "Objective perceptual multimedia video quality measurement in the presence of a full reference," August 2008, https://www.itu.int/rec/T-REC-J.247/en (Accessed 15/1/2021)

[S5] List of PEVQ licensees (list is continuously updated) http://www.pevq.com/licensees.html (Accessed 16/1/2021)

[S6] ITU-T Standard P.563 "Single-ended method for objective speech quality assessment in narrow-band telephony applications," May 2004 https://www.itu.int/rec/T-REC-P.563/en (Accessed 15/1/2021)

[S7] Corroboration from Assistant Vice President, Product Management, Netscout Systems. Confirming that the work from the University of Essex has been incorporated into NetScout nGeniusOne leading to global use and a substantial revenue stream for NetScout.

[S8] InfoTrack for Unified Communications, "Impact of Microsoft Skype for Business on the Enterprise Voice Market—2016," September 2016, https://news.microsoft.com/uploads/2017/01/IUC-MS-Skype4B-2016-Full-Report-10-21-16.pdf (Accessed 15/1/2021)

[S9] Solution Brief from Netscout "Managing Voice, Video Call Quality Issues in Contact Centers with the nGeniusONE Service Assurance Platform," 2019, https://www.netscout.com/sites/default/files/2019-09/ESB_023_EN-1901%20-%20Managing%20Voice%20Video%20Call%20Quality%20Issues.pdf (Accessed 11/09/2020)