Monitoring Scenario for Non-Intrusive Quality Assessment of Teleconference Systems

Leonardo O. Nunes, Flávio R. Ávila, Luiz W. P. Biscainho, Bowon Lee, Amir Said, Ronald W. Schafer

***Resumo*— A demanda por alta qualidade de experiência em sistemas de comunicação tem aumentado com a utilização de sistemas de teleconferência, o que, por sua vez, estimulou a procura por sistemas confiáveis de avaliação de qualidade. Ferramentas para monitoração durante a operação do sistema, em particular, permitem que os sinais sendo transmitidos tenham a sua qualidade avaliada e as degradações que o estão atingindo identificadas sem interrupção do serviço. Neste artigo, são apresentados dois cenários para monitoração de qualidade de sistemas de teleconferência, cada um contendo dois pontos de medição onde podem ser posicionadas ferramentas de monitoração não-intrusiva (INMDs). São levantados os tipos de degradação que podem ser monitorados em cada ponto de medição, além de critérios para o projeto dos respectivos INMDs. Os modelos propostos podem servir de guia para a geração de sinais degradados adequados para testes subjetivos.**

***Palavras-Chave*— teleconferência, degradações acústicas, avaliação não-intrusiva de qualidade, INMD**

***Abstract*— Teleconference systems have increased the demand for high quality of experience in speech communication, which, in turn, calls for reliable quality assessment tools. In-service monitoring tools, in particular, aim to assess the quality of the transmitted signal as well as to identify impairments that might occur during the operation of the system. This paper presents two scenarios for quality monitoring of teleconference systems, each of them containing two measurement points where in-service non-intrusive measurement devices (INMDs) can be located. The impairment types that can be monitored at each measurement point are elicited, as well as design constraints on their respective INMDs. The proposed models can also guide the generation of impaired signals to be employed in subjective listening tests.**

***Keywords*— teleconference, acoustic degradations, non-intrusive quality assessment, INMD**

I. INTRODUCTION

The ubiquitous presence of speech communications has led to an increase in demand for quality of experience [14]. Modern teleconference systems, in particular, call for dedicated quality assessment tools that guarantee the satisfaction of their users' high expectation.

The most reliable method for measuring the quality of a speech communication system is through subjective listening tests. In such tests, the quality of a given signal is measured by asking a group of subjects to grade it in a given scale, and averaging the attributed grades so as to obtain a single mean opinion score (MOS). Since the set-up and execution of subjective listening tests is time-consuming and expensive [3], [8], [2], computational (so-called objective) methods to emulate them gained attention of both research and industry communities [13], [6]. These automatic quality assessment (QA) methods can be signal-based [14], such as the PESQ algorithm [13], or parametric [14], such as the E-Model [12].

Another possible way to classify these tools is regarding the stage in which they are to be employed in a communication system. A QA tool can be used during system design, maintenance, or operation[1], each phase requiring different tools with varying specifications.

During the design and maintenance phases, the operator has complete control over the system and can perform as many tests on the system as desired. Intrusive double-ended (i.e., where both degraded and non-degraded signals are available) [14] QA tools can be employed to predict the perceptual quality of the transmission systems [5], [6], [18] and/or to identify possible problems [16]. A set of test points [2] must be properly defined, as briefly discussed in Section II-A, for this task.

Quality assessment of systems during their operation, on the other hand, calls for tools that are able to work non-intrusively [14]. Such monitoring devices [17] must access the transmission system internally, in order to gather the necessary information without interruption.

In ITU standard P.651 [11], different In-Service Non-Intrusive Measurement Devices (INMD) for voice-grade parameters, in both circuit-switched and packed-based telephony networks, are specified. As defined in that document, "The parameters which are accessible via INMD measurements can only be obtained from the signals (of one direction or of both directions) at one specific point in the network.", hence each INMD is located at a single point at the network and can be double-ended or single-ended (i.e., where only the degraded signal is available to the INMD) [14].

This paper describes a general model for teleconference systems including measurement points where INMDs can be connected, and elicits degradations that can be identified/assessed at each location. Ideally, an INMD located at a given measurement point could house a signal-based [15] QA tool or a degradation type classifier [16] responsible for monitoring the quality of the telepresence system during its operation.

The structure of the paper is as follows. In Section II, the teleconference model originally presented in [2] is briefly described. The new non-intrusive scenario for teleconference

Leonardo O. Nunes, Flavio R. Ávila, and Luiz W. P. Biscainho are with the Signal Processing Laboratory, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. E-mail: ({lonnes, flavio, wagner}@lps.ufrj.br)

Bowon Lee, Amir Said, and Ronald W. Schafer are with the Mobile and Immersive Experience Lab in HP Laboratories, Palo Alto, CA, USA. E-mail: ({bowon.lee, amir_said, ron.schafer}@hp.com)

[1] A similar categorization is performed in [15].

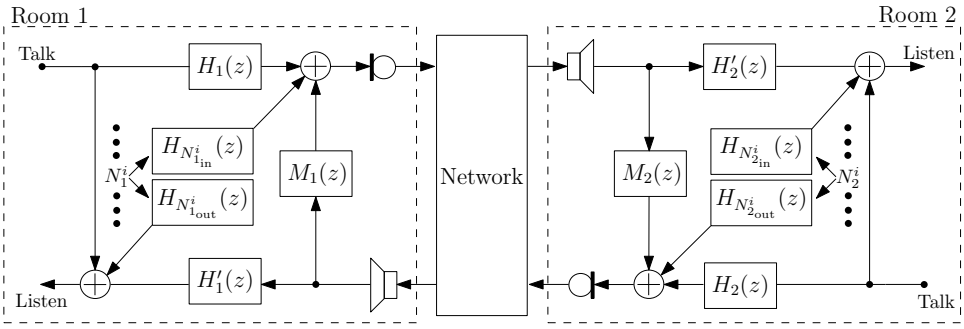

Fig. 1. Complete teleconference model. Figure adapted from [2].

systems is presented in Section III. In Section IV, the conclusions are drawn.

II. Teleconference Model

In this section, the model used as a starting point for the creation of both intrusive and non-intrusive scenarios is reviewed. This model was originally presented in [2] and is restricted to a conversation between two participants.

An overall model for a teleconference system is shown in Figure 1. In that figure, one can see 1 in Room 1 in conversation with 2 in Room 2. Voice coming out of the mouth of 1 reaches his/her own ears (through direct path only, for the sake of simplicity) and microphone 1 through Room 1 response $H_1(z)$. Each noise signal N_1^i from a given source (such as an air-conditioner or a computer) i inside the room, after being modified by and along their respective paths, reaches the talker's ears and microphone 1 also. On the other hand, voice coming from Room 2 through loudspeaker in Room 1 is directed to the ears of Talker 1 through room response $H_1'(z)$, and is acoustically fed back to microphone via $M_1(z)$. An analogous description suits Room 2. Network-induced degradations such as delay, attenuation, packet loss, and jitter are implicitly depicted inside the grey box.

In [2], two scenarios for intrusive quality assessment of teleconference systems based on the model shown in Figure 1 were described. These two scenarios are briefly presented in the next section.

A. Intrusive QA Scenario

The objective of the intrusive scenarios is twofold [2]: (1) representing the targeted impairments in a way that allows the controlled generation of a comprehensive database of degraded signals to aid in the development of automatic QA tools; (2) allowing the definition of access points to reference signals as well as signals under test that are meaningful for both subjective and objective tests.

The model depicted in Figure 1 can be simplified to allow the independent characterization of echo, noise and reverberation effects in each room. This can be done by separating the general model in two parts: local scenario and remote scenario.

The so-called local scenario is shown in Figure 2. It can characterize degradations acoustically generated at the talker's side, i.e. background noise originated by local sources ($\overline{N}_1$) and reverberation due to the local room impulse response (RIR).

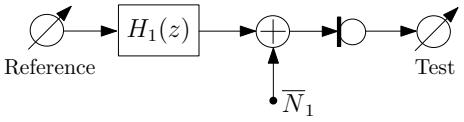

Fig. 2. Local scenario model [2].

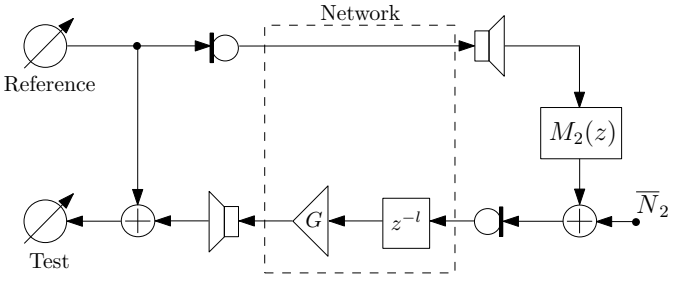

Fig. 3. Remote scenario model [2].

The so-called remote scenario is shown in Figure 3. It can characterize degradations acoustically generated at the talker's opposite side, i.e. mainly acoustic echo resulting from loudspeaker-microphone coupling combined with transmission paths' delay. Local effects are discarded, as if $H_1(z) = 1$ and $\overline{N}_1 = 0$ in Figure 2. Possible network distortions are not considered, and the effect of the signal transmission is summarized by delay l and gain G. Additive noise $\overline{N}_2$ and room response $M_2(z)$ make for a more realistic modeling of the remote part.

In both cases, the multiple noise sources have been encapsulated into a single one ($\overline{N}_1$ or $\overline{N}_2$) with

$$\overline{N}_1(z) = \sum_{i=0}^{M_1-1} H_{N_{1_{\text{in}}}^i}(z) N_1^i(z) \tag{1}$$

and

$$\overline{N}_2(z) = \sum_{i=0}^{M_2-1} H_{N_{2_{\text{in}}}^i}(z) N_1^i(z), \tag{2}$$

where M_1 and M_2 are the numbers of noise sources in Room 1 and Room 2, respectively.

A description of how these two scenarios can be used in intrusive QA tests can be found in [2].

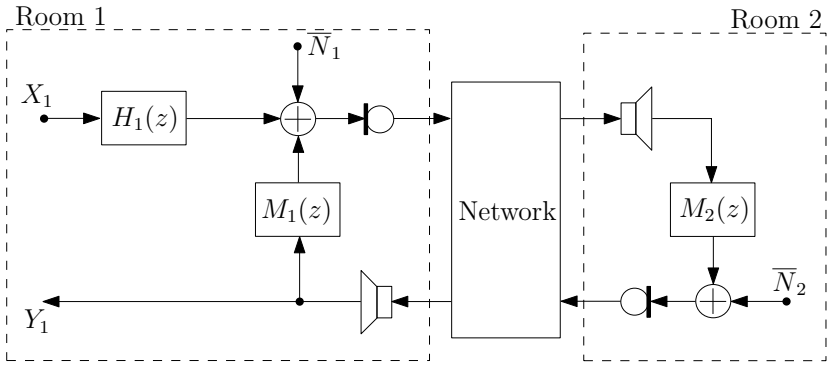

Fig. 4. Simplified teleconference model. Only the transmission/reception cycle for talker 1 is considered.

The two models presented in this section are unsuitable for non-intrusive QA because: (1) their signal capture points are not readily available for in-service acquisition; (2) they focus in acoustically-induced degradations, which is a reasonable constraint for network testing when different degradation sources can be de-coupled but is not the case for online measurement. In the next section, two new models specifically devised for online monitoring of teleconference systems are described.

III. Monitoring Scenario

In this section, two new scenarios for teleconference systems are presented. Their target is to provide meaningful measurement points for non-intrusive quality monitoring of a teleconference system. Even though these test points have been intended for objective quality measurements, their possible use for subjective listening tests is also discussed.

Once more, the starting point is the overall teleconference model depicted in Figure 1, from which the simplified model shown in Figure 4 is derived. In the latter:

1) Only the transmission/reception cycle related to the voice of talker 1 is considered.
2) A single additive noise signal as a combination of several noise sources is considered inside each room.
3) Room response $H_{N_j}(z)$ represents the path from the single noise source N_j to the microphone in Room j, with $j \in \{1, 2\}$.
4) The received signal is taken at the loudspeaker output, hence the room response from the loudspeaker to the talker's ears is disregarded.

Item 1 above implies that the monitoring strategies considered henceforth operate only when just one talker is active, i.e. there is no double-talk. In practice, this would require the use of a double-talk detector [9] in tandem with the quality monitoring tools. A similar simplification as the one described in Item 2 was also made in Section II-A [2], and was found to be a reasonable compromise between the simplicity of the model and its generality. The asumption in Item 4 regards the fact that a monitoring system has only access to the signal up to the loudspeaker output, and that possible impairments occurring between the loudspeaker and the subject's ears are unobservable from a monitoring point of view.

In the next two sections, the model depicted in Figure 4 is further divided into two scenarios: one responsible for degradations that might occur due to the direct path between Room 1 and Room 2; and the other responsible for the degradations that might occur due to the feedback path from/to Room 1. Of course, both models can be side-reversed.

A. Direct Path Model

In this scenario, shown in Figure 5, degradations that occur through the transmission system when there is no feedback between the two rooms are considered. This can happen if the Acoustic Echo Cancelers at both ends are working properly or if headphones are used in Room 2 instead of loudspeakers.

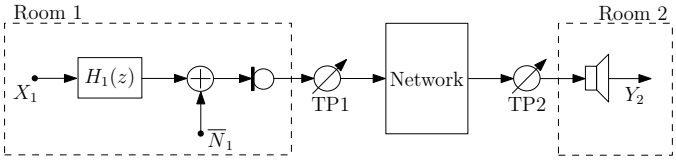

Fig. 5. Direct path scenario. TP in the diagram stands for "test point".

Two test points, TP1 and TP2, are identified in this scenario. The first allows the evaluation of the signal that is going to be transmitted to Room 2, whereas the second allows the evaluation of the signal received in Room 2. Below, the degradations that can be detected at each test point are described.

- TP1: acoustic degradations that might occur in Room 1 (reverberation and background noise), plus possible non-linear distortions caused by the audio capturing device (microphone and A/D converter), such as magnitude clipping.
- TP2: all degradations at TP1 plus network distortions, such as coding artifacts, jitter, and signal gaps.

TP1 in the Direct Path Model is similar to the test point of the Local Scenario described in Section II-A, in the sense that both consider almost the same degradations. On the other hand, TP2 is a more encompassing evaluation point, since it allows the inclusion of network-induced impairments, and thus estimating the overall perceptual quality for a listener located at Room 2 when acoustic degradations originated inside Room

2 are disregarded[2]. Both test points are single-ended, i.e. they only have access to the degraded signal.

As regards subjective listening tests, for TP1, Degradation Category Rating [10] tests similar to the ones suggested in [2] can be performed, as long as the reference signal is also captured, which can be done the same way as in the Local Scenario. Alternatively, Absolute Category Rating [10] (ACR) tests can be carried out by submitting to evaluation only the signal observed at this test point. ACR tests are the most adequate for application to the signals detected at TP2.

It should be noted that the number of combinations of impairments that may occur at TP2 can make the design of monitoring tools too challenging a task. In any case, the degradation decomposition approach proposed in [19], [20] can be applied at each test point. This way, quality evaluation is carried out over three proposed quality dimensions [20], making the monitoring system more robust to impairment type combinations. A similar approach is being considered for the new ITU standard P.863 for full-reference quality evaluation of speech signals [7]. Moreover, subjective listening tests can also be carried out to assess each quality dimension separately [4], by considering only those impairments associated with the dimension under test.

B. Feedback Path Model

In this scenario, depicted in Figure 6, only those degradations related to the feedback path caused by coupling between microphone and loudspeakers inside Room 2 are considered. Differently from the direct path model, now the acoustic echo canceler (AEC) at room 2 is explicitly shown. The degradations occurring inside Room 1 are not taken into account, since they can be evaluated via the direct path scenario. Moreover, the acoustic coupling that happens in Room 1 (due to $M_1(z)$) is also neglected for two reasons: (1) the AEC at Room 1 is assumed to be working properly; (2) a high signal attenuation is expected due to the double feedback loop. Two test points

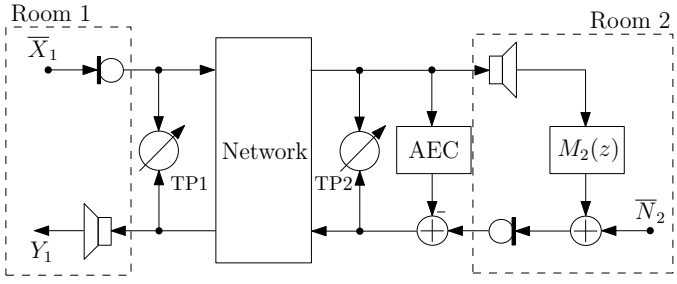

Fig. 6. Feedback path model. TP in the diagram stands for "test point"

were envisaged for this model:

- TP1: that allows the evaluation of the acoustic echo and noise originated in Room 2.
- TP2: that allows the evaluation of the AEC located at Room 2.

Both test points are double-ended, having access to both reference and degraded signals. TP1 in particular is closely related to the Remote Scenario (see Section II-A) if the network is simplified to only gain and delay. TP2, on the other hand, is of particular interest if one needs to isolate the effect of the AEC, in order to monitor its performance.

TP1 cannot be used to directly provide signals for subjective listening tests. This happens because echo is perceived only when the feedback signal Y_1 is acoustically summed with the signal $\overline{X}_1$. Hence, in order to perform subjective listening tests for TP1, the degraded signal actually evaluated should be $\overline{X}_1 + Y_1$. In this case, as mentioned before, both the reference and degraded signals of TP1 are exactly the same as the ones defined for the Remote Scenario, assuming that the effect of the AEC is simply an attenuation of the feedback speech signal. As with TP1, the signals at TP2 need to be conformed before being used in listening tests. Alternatively, TP1 can be also used in talking-quality[3] subjective tests, such as the ones used in [1]; in this case the model can be used to generate the degraded signal in real-time, while the subject speaks.

No degradation decomposition in terms of those described in [20], [19] can be made at any test point in this scenario, since echo-degraded signals have not been considered in those works' experiments.

IV. Conclusion

In this work, two new scenarios for non-intrusive monitoring of teleconference systems were described, related to the direct and the feedback paths. For each scenario, test points (TP) suitable for degradation type identification and quality assessment were proposed. For the case of the feedback path model, it was found that double-ended non-intrusive measurement points could be used for detecting and assessing the system as regards acoustic echo. For the direct path model, single-ended measurement points can be used to assess acoustic degradations induced by the room where the talker is located as well as possible network distortions. These assessment points allow the development of dedicated tools for online quality monitoring of teleconference systems. Also, the design of subjective listening tests that can be employed as ground-truth for this kind of QA tools can be guided by knowing which degradations are accessible at each measurement point, as described in this paper.

Acknowledgment

Dr. Biscainho wishes to thank fomenting agencies CNPq and FAPERJ for financial support of his research projects. Leonardo O. Nunes would like to thank CNPq for partially supporting his work. This R&D project is a cooperation between Hewlett-Packard Brasil Ltda. and COPPE/UFRJ, being supported with resources of Informatics Law (no 8.248, from 1991).

References

[1] R. Appel and J. G. Beerends. On the quality of hearing one's own voice. *Journal of the Audio Engineering Society*, 50(4):237–248, April 2002.

[2] The latter could be measured by a reversed TP1.

[3] When a subject is asked to evaluate the quality of hearing his/her own voice.

[2] F. R. Ávila, L. W. P. Biscainho, L. O. Nunes, A. F. Tygel, B. Lee, A. Said, T. Kalker, and R. W. Schafer. A teleconference model with acoustic impairments suitable for speech quality assessment. In *Anais do XXVII Simpósio Brasileiro de Telecomunicações*, number 4-58318, Blumenau, Brazil, October 2009. SBrT.

[3] S. Bech and N. Zacharov. *Perceptual Audio Evaluation – Theory, Method and Application*. Wiley, Chichester, England, 2007.

[4] J. G. Beerends, B. Busz, P. Oudshoorn, J. Van Gugt, K. Ahmed, and O. Niamut. Degradation decomposition of the perceived quality of speech signals on the basis of a perceptual modeling approach. *Journal of the Audio Engineering Society*, 55(12):1059–1076, December 2007.

[5] L. W. P. Biscainho, P. A. A. Esquef, F. P. Freeland, L. O. Nunes, A. F. Tygel, B. Lee, A. Said, T. Kalker, and R. W. Schafer. An objective method for quality assessment of ultra-wideband speech corrupted by echo. In *Proceedings of the International Workshop on Multimedia Signal Processing*, number 149, Rio de Janeiro, Brazil, October 2009. IEEE.

[6] B. C. Bispo, P. A. A. Esquef, L. W. P. Biscainho, A. A. de Lima, F. P. Freeland, R. A. de Jesus, A. Said, B. Lee, R. W. Schafer, and T. Kalker. EW-PESQ: A quality assessment method for speech signals sampled at 48 kHz. *Journal of the Audio Engineering Society*, 58(4):251–268, April 2010.

[7] N. Côté, V. Gautier-Turbin, and S. Möller. Influence of loudness level on the overall quality of transmitted speech. In *123rd AES Convention*, number 7175, New York, USA, October 2007.

[8] T. H. Falk. *Blind Estimation of Perceptual Quality for Modern Speech Communications*. Phd thesis, Department of Electrical and Computing Engineering, Queen's University, Kingston, Canada, December 2008.

[9] ITU-T. *Rec. P.56: Objective Measurement of Active Speech Level*. International Telecommunication Union, Geneva, Switzerland, 1993.

[10] ITU-T. *Rec. P.800: Methods of Subjective Determination of Transmission Quality*. International Telecommunication Union, Geneva, Switzerland, 1996.

[11] ITU-T. *Rec. P.561: In-service non-intrusive measurement device – Voice service measurements*. International Telecommunication Union, Geneva, Switzerland, 2002.

[12] ITU-T. *Rec. G.107: The E-model, A Computational Model for Use in Transmission Planning*. International Telecommunication Union, Geneva, Switzerland, 2005.

[13] ITU-T. *Rec. P.862: Perceptual Evaluation of Speech Quality (PESQ): Objective Method for End-to-end Speech Quality Assessment of Narrow Band Telephone Networks and Speech Codecs*. International Telecommunication Union, Geneva, Switzerland, 2005.

[14] ITU-T. *Rec. P.10 & G.100 Amd. 2: Vocabulary and effects of transmission parameters on customer opinion of transmission quality*. International Telecommunication Union, Geneva, Switzerland, 2008.

[15] S. Möller and A. Raake. Telephone speech quality prediction: Towards network planning and monitoring models for modern network scenarios. *Speech Communication*, 38(1):47–75, January 2002.

[16] L. O. Nunes, F. R. Ávila, L. W. P. Biscainho, B. Lee, A. Said, T. Kalker, and R. W. Schafer. Degradation type classifier for full band speech contaminated with echo, broadband noise, and reverberation. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011. To appear.

[17] A. Raake. *Speech Quality of VoIP – Assessment and Prediction*. Wiley, Chichester, England, 2006.

[18] A. Rix, M. Hollier, A. Hekstra, and J. G. Beerends. Perceptual evaluation of speech quality (PESQ), the new ITU standard for end-to-end speech quality assessment, Part I – Time-Delay Compensation. *Journal of Audio Engineering Society*, 50(10):755–764, October 2002.

[19] M. Wältermann, A. Raake, and S. Möller. Modeling of integral quality based on perceptual dimensions – a framework for a new instrumental speech-quality measure. In *Proceedings of the 8th Fachtagung Sprachkommunikation*, pp. 1–4, Aachen, Germany, October 2008. ITG.

[20] M. Wältermann, K. Scholz, A. Raake, U. Heute, and S. Möller. Underlying quality dimensions of modern telephone connections. In *Proceedings of the International Conference on Spoken Language Processing – Interspeech*, pp. 2170–2173, Pittsburgh, USA, September 2006. ISCA.