IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 9, 2021, accepted March 15, 2021, date of publication March 19, 2021, date of current version March 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3067459

Assessment of Extreme Communication Environment With Ultralow SNR: A Benchmark

LEI ZHANG[1], (Member, IEEE), XILE ZHAO[1], AND XIN LI[1,2]
[1]College of Computer Science, Guangdong University of Petrochemical Technology, Maoming 525000, China
[2]Gaitech Robotics Inc., Centrical Mechatronics Industrial Park, Zoucheng 273500, China

Corresponding author: Xin Li (lixin@gdupt.edu.cn)

This work was supported in part by the Project of Educational Commission of Guangdong Province of China under Grant 2018KCXTD019, in part by the Science and Technology of Guangdong Province under Grant 191103104554996, and in part by the Key Projects of Shandong Natural Science Foundation under Grant ZR2020KF022.

ABSTRACT Accurate estimation of subjective assessment plays an essential role in not only speech quality perception, but also communication environment assessment. Traditional speech quality perception is almost always with an environment, in which the content of speech can be heard clearly. Unlike speech quality assessment, in extreme communication environment with ultralow SNR, like short wave channel with active jamming, speech intelligibility is impaired. Under this condition, subjective assessment as absolute category ranking (ACR) whose scores are made by experienced staffs, and many objective measurements cannot handle this situation. In this paper, we propose information damage level (IDL) to replace subjective ACR as subjective assessment. IDL is the average of scores which is marked on the subjective recognition rate. Under extreme communication environment, it can effectively avoid excessive differences from people to people to some extent. We also provide a new dataset collected in an environment with active jamming, whose speech files are recoded under three different environments, named as indoor simulation environment (EN1), outdoor simulation environment (EN2), and outdoor real environment (EN3). We also benchmark a novel open framework on random forest for direct predicting subjective assessment by combining all possible objective measurements. Experiment results prove the effectiveness of our open framework together with our dataset.

INDEX TERMS Ultralow SNR, information damage level, random forest, direct subjective assessment.

I. INTRODUCTION

Communication environment assessment by speech signal is a problem with a long history. It can be applied in two aspects. One is to measure the quality of communication system to improve the speech intelligibility, such as telecommunication channel. While the other one is for predicting the interference intensity to prevent the normal communication, especially in military applications.

With the development of speech signal processing, speech transmission quality assessment provides an effective way to evaluate communication environment by comparing speech signals between the sender and the receiver, or distinguishing speech intelligibility only based on the receiver speech signal. [1] presents an objective intelligibility measure based on time-frequency (TF) weighting approach under noisy condition. [2] adopts this objective intelligibility measure into motoring speech disorders evaluation. [3], [4] proposes spectral subspace analysis on principal component analysis and approximate joint digitalization to assess speech intelligibility. [5] considers the auditory properties from the attention mechanism to model the mobile audio objective quality assessment. [6], [7] introduces a perceptual, binaural audio-quality model to perceive spatial quality differences between two audio signals in multi-channel reproduction. More recently, [8] uses convolutional neural network (CNN) to estimate the per-frame quality and adopt recurrent neural network (RNN) to aggregate the per-frame values over time, to estimate the overall speech quality. [9] predicts speech quality by a model based on the outputs of an automatic speech recognizer, and in [10], a model based on a BiLSTM network is shown to assess speech quality.

Although some effective efforts were made in past years, communication environment assessment is still an open

The associate editor coordinating the review of this manuscript and approving it for publication was Jing Liang.

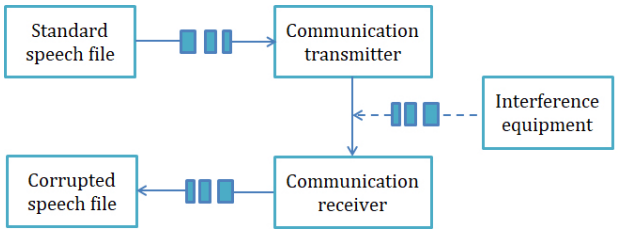


FIGURE 1. Equipment connection in data collection.

problem and many problems are worthy to research, especially for some severe conditions with negative signal noise ratio (SNR). When interference in communication environment seriously affect the content understanding, traditional absolute category ranking (ACR) measurement is difficult to assess the subjective sensation. Another problem is the gap between subjective and objective measurement. Most proposed approaches only focus on objective measurement, which is not intuitive to user. Additionally, there is no publicly available standard database which is essence to spur interest and inspire novel ideas.

In fact, environment assessment on speech signal has two ways to evaluate the system efficiency, i.e., subjective assessment from human being and objective assessment from machine. Unlike some approaches which focus only on objective assessment, in this paper, we propose a unified model that can map diverse objective measurements to subjective assessment, whose final results are similar as the score marked by human being. As a consequence, the major contributions of this paper are as follows:

- By a series of listening experiments of experts, a new subjective assessment named as information damage level (IDL) is put forward, which is much more suitable for ultralow SNR environment assessment.
- By proposed open framework which can be expanded without further burden, our approach can directly assess the quality of communication environment with intuitive results. Our benchmark has formed as a military standard file (GJB 4405B-2017) in China.

II. INFORMATION DAMAGE LEVEL

Unlike normal communication system, in some special applications, an interference equipment is used to destroy the communication channel, which leads to the bad quality of speech. Our paper aims to handle the most severe situation, and our data acquisition system is shown in Fig. 1. Besides transmitter and receiver which are far away from each other, in our system there is an interference equipments placed in the third position. We mainly analysis FM noise jamming interference in this paper.

A. CORRUPTED SPEECH BY INTERFERENCE EQUIPMENT

Fig. 2 gives the comparison of spectrum between clean and corrupted speeches with short wave communication channel interfered by FM noise jamming signal. It can be seen from spectrum in Fig. 2 (b) that clean speech is composed by 15 groups, with 4 characters in each group. Since speech signal possesses the properties of formant, fundamental frequency and harmonic frequency, it displays the horizontal stripes in each bar of Fig. 2 (b). However, all these behaviors almost disappear in Fig. 2 (a) due to the active jamming, which also result in the bad intelligibility of speech. If we adopt traditional absolute category ranking (ACR) as the human judged score on a scale of 1 (bad) to 5 (excellent), most data may be categorized into grade 1. Additionally, it can be seen that the start points of (a) and (b) in Fig. 2 are different because of communication delay, coding algorithm delay and so on. In order to align the speeches at both ends, a synchronous head which is much more robust to active jamming is added, consisting of two alternative single frequency signals.

Summarily, we can conclude three issues from the comparison:

- We could no longer use traditional subjective assessment as ACR to label the data's grade in this special case.
- Synchronous head is also degraded, and with serious condition, only part of the synchronous head is left.
- The delay is kind of random, and with no possible to control beforehand.

B. INFORMATION DAMAGE LEVEL

ACR scale is mainly for speech quality subjective assessment, in which the most contents of speech are understandable. However, under extreme communication environment with ultralow SNR, since most contents of receiver file are blurred, it is hard to distinguish the difference between fair and good, or poor and bad. However, for ACR, when human judges the speech quality, he has a freedom to give a score from 1 (bad) to 5 (excellent) depending on his feeling to the speech. It will increase the risk of wrong grade results, such as misclassifying level 4 to level 5. Although mean opinion scores (MOS) on ACR can eliminate the error by meaning operation to some extent, there is still some bias.

In order to give more precise grade according to subjective sensation, we name a new concept as subjective recognition rate (SRR) and relate the quality assessment to SRR. SRR means the recognition rate from human being when listening to a speech record. We believe that SRR can reflect the intelligibility of speech, which is the essence of speech quality. By a large of aural comprehension experiments on experts which is familiar to both short wave communication channel and FM noise jamming mode, we build the relations between SRR and ACR as Fig. 3.

In this figure, strict restrictions between SRR and ACR are made shown as red crosses. It means that 10% SRR corresponding to level 1, 50% SRR corresponding to level 2, 80% SRR corresponding to level 3, 95% SRR corresponding to level 4 and 100% SRR corresponding to level 5. However, loose limitations are applied between red crosses, as shown as green dashed line or blue real line in Fig. 3, which reflects

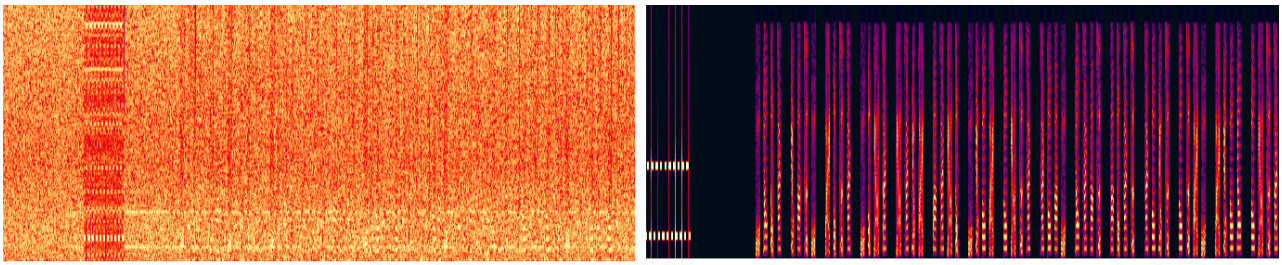


FIGURE 2. Comparison of spectrum between clean and corrupted speeches with short wave communication channel interfered by FM noise jamming signal.

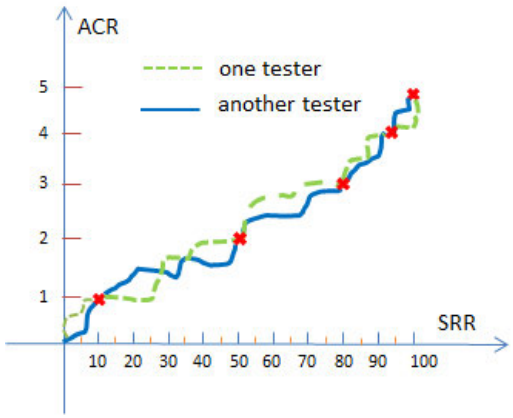


FIGURE 3. Relations between SRR and ACR.

TABLE 1. Relationship between ACR and information damage level.

ACR		SRR	IDL
5	Excellent	$100\% \geq r > 99\%$ $99\% \geq r > 97\%$ $97\% \geq r > 95\%$	$1.3 > G_i \geq 1.0$ $1.7 > G_i \geq 1.3$ $2.0 > G_i \geq 1.7$
4	Good	$95\% \geq r > 90\%$ $90\% \geq r > 85\%$ $85\% \geq r > 80\%$	$2.3 > G_i \geq 2.0$ $2.7 > G_i \geq 2.3$ $3.0 > G_i \geq 2.7$
3	Fair	$80\% \geq r > 70\%$ $70\% \geq r > 60\%$ $60\% \geq r > 50\%$	$3.3 > G_i \geq 3.0$ $3.7 > G_i \geq 3.3$ $4.0 > G_i \geq 4.7$
2	Poor	$50\% \geq r > 40\%$ $40\% \geq r > 25\%$ $25\% \geq r > 10\%$	$4.3 > G_i \geq 4.0$ $4.7 > G_i \geq 4.3$ $5.0 > G_i \geq 4.7$
1	Bad	$10\% \geq r > 0\%$	$G_i \geq 5.0$

the subjective feeling of speech quality from different testers. It means that unlike ACR which the freedom is from 1 to 5 for listener, the freedom is limited into a controlling range (e.g., range between red crosses) when a listener give a subjective evaluation.

We further propose an evaluation rule named as information damage level (IDL) G_i by more detailed division on SRR r listed in table 1. It is worth to note that IDL inverse to ACR, which is more convenient to reflect the damage of speech quality or the effectiveness of interference. Similar as ACR, average is conducted for information damage level from expert listeners to eliminate human tendency.

In fact, proposed IDL is kind of combination of objective and subjective assessment. It gives tester some freedom to score within level, which avoids the major errors across grades.

III. ASSESSMENT METHODOLOGY

Our paper attempts to approximate human experience scoring the quality of environment by speech. In order to directly evaluate subjective assessment according to human sensation, we propose an open framework based on random forest model, which can combine different objective measures and select effective ones mapping to subjective measure during learning stage. The whole framework is shown in Fig. 4.

From Fig. 4, it can be seen that our open framework is divided into time alignment, speech representation, objective measurement and random forest. We introduce the details in following subsections.

A. TIME ALIGNMENT

As shown in Fig. 2 (b), a synchronous head composed of two alternative single frequency signals is added in sender end before transmitting normal signal. By synchronous head, we aim to align two speech signals at sender and receiver ends. Fig. 5 provides different conditions of synchronous head at receiver. It can be seen that for some interference, the synchronous head could also be damaged seriously, such as condition 2 in Fig. 5 (b), where only a small part of synchronous head are survived from heavy interference.

From Fig. 5, it can also be drawn that, although the synchronous head is also damaged, the single frequency signal in the synchronous head still keeps the maximum frequency energy among all signals. Thus we can judge the real start point of the speech by the location of last maximum frequency intensity, as the green line in Fig. 5.

B. SPEECH REPRESENTATION

Since our mapping model, random forest is with selection property, in speech representation, the whole framework

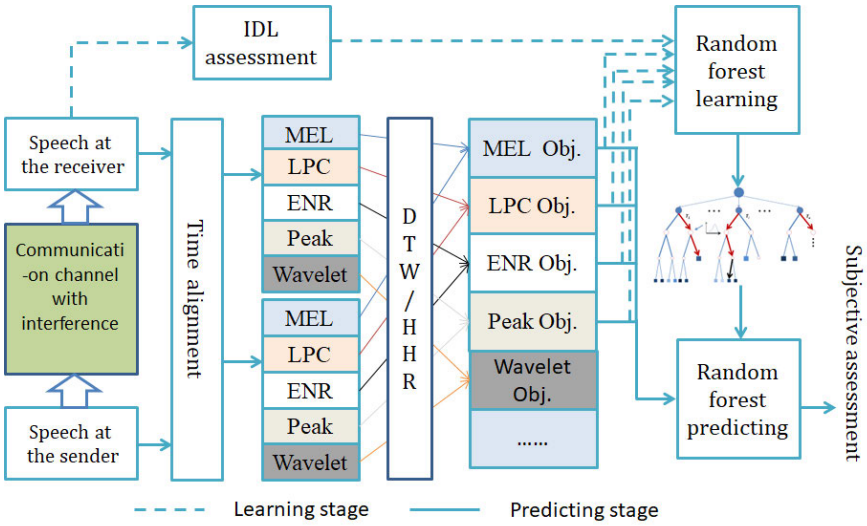


FIGURE 4. The whole system framework, where the dashed line to random forest is involved in learning stage.

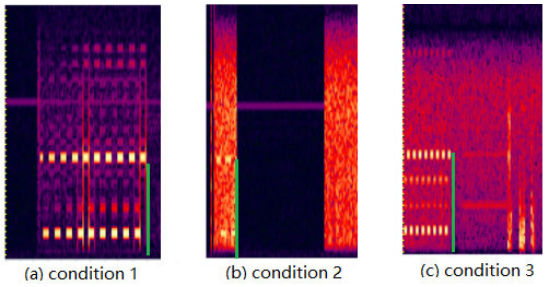


FIGURE 5. Synchronous head at receiver under different conditions.

keeps open state. It means that we can extract more representations as long as the delay within users' toleration. As shown in Fig. 4, speech representation such as Mel energy, linear prediction coefficient(LPC), ENR (engery rate), formant peak, and wavelet coefficient, can be extracted here from different aspects, such as time-frequency analysis and human auditory. Table 2 lists the detailed character of each representation.

- MFCC and Mel energy

Besides removing DC component, static Mel-Frequency cepstral coefficients(MFCC) after cepstral lifting plus delta coefficients composes 24 order MFCC vector. In addition, cepstral mean normalization (CMN) is added here to depress the noise influence. Mel energy is the same procedure in MFCC except the cosine transform and cepstral lifting.

- LPC and LPCC

Different from MFCC or Mel energy which models the filter effect of basement membrane in cochlea, LPC and linear predictive cepstral coding(LPCC) characterize the vocal tract variation during speaking. Normal LPC is extracted by Durbin algorithm with 12-pole filter. As for LPCC, cepstral lifting and CMN are applied which are similar as those in MFCC.

- Wavelet feature

DB4 wavelet is selected here to decompose one frame speech into four levels wavelet coefficients. For each level, the energy, mean, variance, maximum and the location of maximum are extracted to describe the speech characters.

- ENR (engery rate)

Each frame energy ratio between sender and receiver is computed and the statistical characteristics as mean and variance can be the measurement.

- Formant peak

Suppose $S(f, n)$ be the spectrum of frame n in f channel by FFT transform. Here, we only focus on the interested frequency between 200∼3500Hz and the energies of spectrum beyond this range are set to a fixed value. Moreover, S_{max} corresponds to the maximum of spectrum over f and n, while S_{mean} is the mean energy of spectrum. According to S_{max}, we normalize $S(f, n)$ as

$$\hat{S}(f, n) = max(S(f, n), S_{max}/\alpha) - S_{mean} \quad (1)$$

where α is determined by pre-analysis of energies of corrupted speeches in experiments, here we select $\alpha = 1e6$ in following experiment. After a high pass filter G_{hp}, $\hat{S}(f, n)$ can be further smoothed as $S'(f, n)$. Since the spectrum is easy disturbed, in order to find the robust spectrum peak, we conduct filter smooth on the differential of $\hat{S}(f, n)$ as below.

$$S'(f, n) = \hat{S}(f, n) - \hat{S}(f, n-1) + 0.98S'(f, n-1) \quad (2)$$

C. OBJECTIVE MEASUREMENT

Objective measurement of above features can be computed on dynamic time wrap (DTW) or HHR, as shown in table 2.

Algorithm 1 Algorithm for HHR

Require:
Formant peak matrix $S_1'(f,n)$ and $S_2'(f,n)$ from clean and corrupted speech, considering frequency range $[f_{min}, f_{max}]$, time lag t_l, frequency lag f_l and differential energy threshold S_T

Ensure:
Hash hit rate between two sequences.

1: **Initialization:** Set initial threshold T as the maximum formant peak among the first 10 frames and all frequency, which is as
$\max_f (\max_{10<n<1} S'(f,n))$;
2: **for** $* = 1, 2$ **do**
3: **for** $n = 1, 2, \ldots, N$ **do**
4: if $S_*'(f,n) < T$, then $S_*'(f,n) = 0$, where $f \in [f_{min}, f_{max}]$
5: $f_*^m = \arg\max_f S_*'(f,n)$
6: $T = \max(T, \sum_{i=1}^{5} w_i top(i))$, where $top(1), \ldots, top(5)$ are the top 5 formant energies among $S_*'(f,n)$ and w_i obeys Gaussian distribution.
7: remember record as $[n : f_*^m(n), S_*'(f_*^m(n), n)]$;
8: **end for**
9: **end for**
10: **for** $n = 1, 2, \ldots, N$ **do**
11: In sequence range $[\max(0, n - t_l), n + t_l]$, if $abs(f_1^m(n) - f_2^m(k)) < f_l$ and $abs(S_1'(f_1^m(n), n) - S_2'(f_2^m(n), n)) < S_T$, then remember record as $[n, f_1^m(n), f_2^m(k), abs(k-n)]$.
12: $h_n = h(n, f_1^m(n), f_2^m(k), abs(k-n))$, where h is hash function;
13: **end for**
14: $HHR = mean(h_n)$;

TABLE 2. Characteristics of different representation and objective computation methods.

Representation	Characteristics	Obj. comp.
Mel energy	auditory model	DTW
LPC	vocal tract model	DTW
wavelet coef.	time-frequency analysis	DTW
formant peak	vocal tract model	HHR
ENR	voice and unvoice analysis	-

- DTW

Given the feature sequences of original and jammed signal after synchronous processing, as shown in table 2, for the first three representations, DTW is adopted to obtain the objective measurement since the lengthes of these two sequence may be different even after handcraft synchronous notation.

$$Dis_{DTW} = \min_p \frac{\sum_{n=1}^{N} d(a_{i(n)}, b_{p(i(n))}) W_n}{\sum_{n=1}^{N} W_n} \quad (3)$$

where $p(i(n)) = j(n)$, which is the match point of $i(n)$, and W_n is the weight which is determined aforehand. a_* corresponds

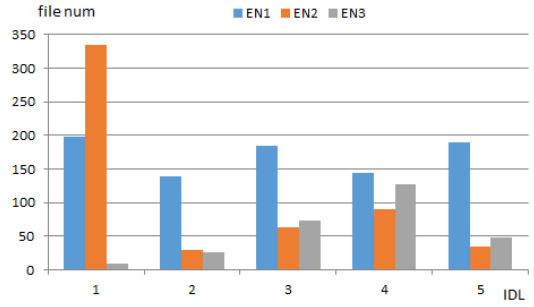


FIGURE 6. Data size with different environment.

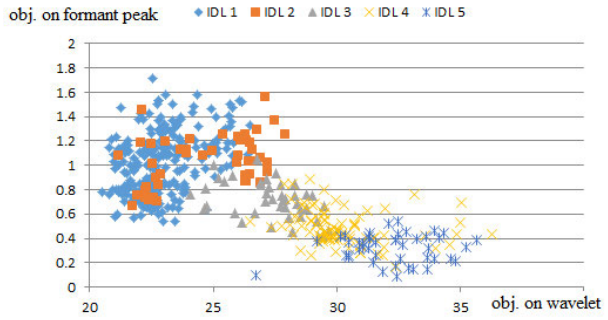


FIGURE 7. Predict error under three environments.

the clean speech representation sequence, while b_* is the corrupted one. N is the clean sequence length.

- HHR

As for formant peak, hash hit rate (HHR) is computed instead of DTW. HHR is listed in algorithm 1. From the algorithm 1, several issues should be noted as

1) computing formant peak

The first two nested **for-end** is to search formant peak. Based on the initial threshold T, we renew a dynamic threshold by adding a Gaussian window weight to smooth n^{th} frame local maximum. We update the threshold with the local maximum of previous frame, thus it can eliminate the nonstationary interference to some extent.

2) matching between clean and corrupt sequences

The third **for-end** is matching procedure. It needs to remember that $S_*'(f,n)$ is a matrix. Because computing of formant peak of corrupt speech may be inaccurate, we loose the matching space as $\{[\max(0, n - t_l), n + t_l], [\max(0, f - f_l), f + f_l]\}$.

3) default parameters

In following experiments, we set default parameters in algorithm 1 as:
frequency range $[f_{min}, f_{max}] = [200, 2500]$
time lag $t_l = 63$
frequency lag $f_l = 30$
differential energy threshold $S_T = \frac{\max T}{10}$.

D. RANDOM FORESTS

Different objective measurements on DTW or HHR can be obtained on the speech representation in table 2. In fact,

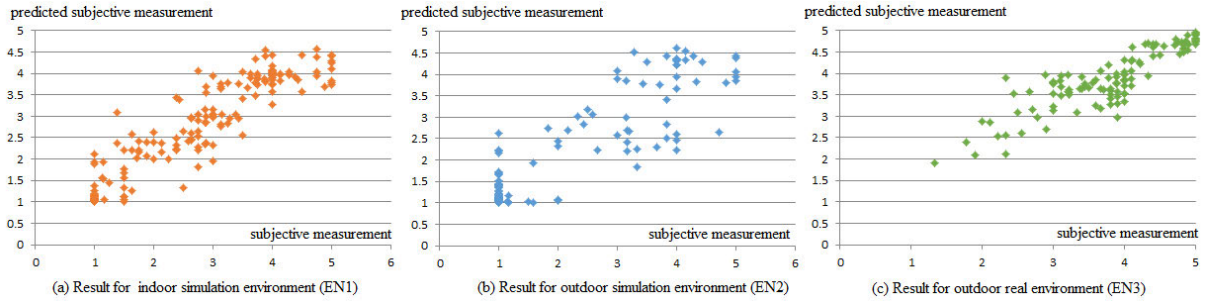


FIGURE 8. Relations between subjective measurement and predict subjective measurement.

corresponding to IDL, those objective measurements are hardly to keep a consistence tendency. Furthermore, ultralow SNR also cause the gap between objective and subjective measurements. Since our goal is to approximate human's experience to make the score of speech quality, our model not only need to bridge the gap between objective and subjective measurements, but also can tolerate the disharmony among objective measurements tendency.

For the big advantages of regress forest such as avoiding overfitting and excellent feature selection ability, it is applied here to adjust the disagreement of objective measurements and fill the gap to subjective measurement.

We hope to consider much more objective measurements from different aspects and build the relations to subjective IDL by random forests. As introduced in [11], random forests undergoes learning and predicting stages as shown in Fig. 4, and it can select and fuse the most representative features closely related to subjective measurement.

By random forests, we can directly build an assessment system imitating human to evaluate environment, which is much more intuitive than objective measurements.

IV. ASSESSMENT RESULT

A. DATA SET

1) COLLECTION CONDITION

The communication system is of short wave communication. Before data collection, we record 10 standard speech files covering 5 men and 5 women speakers. Each file contains 48 characters from number 0 $\sim$ 9, which is played at transmitter and received at receiver.

We configure three kinds of communication environment to collect the receiver data, which named as indoor simulation environment (EN1), outdoor simulation environment (EN2), and outdoor real environment (EN3). In indoor simulation environment, transmitter and receiver are connected by a simulation communication device, which can be injected into interference signal. While for outdoor real environment, transmitter, receiver and interference devices are placed into three cars with antenna on the top, which are 10km far away from each other. Instead of three separate spots in outdoor real environment, in outdoor simulation environment, these three equipments are placed in a open field.

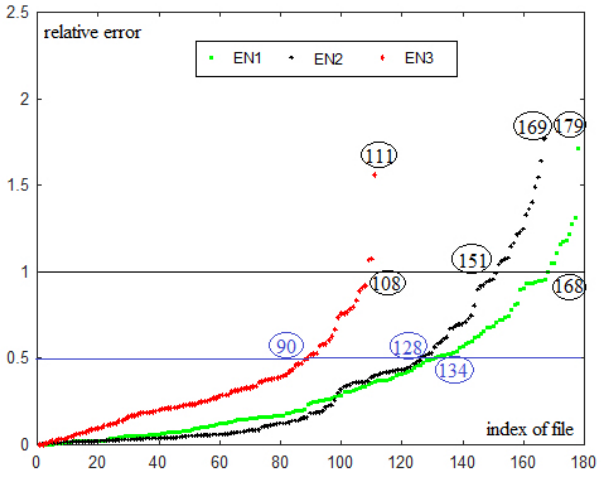


FIGURE 9. Predicted error under three environments.

In order to eliminate communication channel delay, we alternatively add the synchronous head with 500Hz and 2000Hz pure tone signal before normal signals.

2) DATASET STATISTICS

Fig. 6 shows the data size with different environments, where EN1 with 857 files, EN2 with 552 files, and EN3 with 276 files. According to IDL, it can be seen that for En1, data size on each level is balanced since adjusting the intensity of interference is relatively simple in indoor simulation environment. While for EN2 and EN3, this kind of balance is hardly to keep, especially in real environment.

B. ANALYSIS ON OBJECTIVE ASSESSMENT

The performance of any single objective assessment on proposed features (MFCC, Mel, LPC, LPCC, Wavelet, ENR, formant peak) is poor. As shown in [11], Wavelet and formant peak are with better performance. Fig. 7 shows data points of different subjective IDL with Wavelet and formant peak in EN3 environment. It can be seen that even with two obj. assessments of Wavelet and formant peak, it is hard to distinguish IDL 1 and IDL 2. As for IDL 3 $\sim$ 5, although the separability is better than IDL 1 and IDL 2, there are still some overlap area.

It can be concluded that only on several objective assessments, it is hard to distinct different IDL correctly. It is worth to note that our framework is a open one, and we can add another objective assessment at any time, and we suppose that different objective assessments can be complementary. We aim to achieve better performance by leveraging the selecting character in random forest next.

C. RESULT ANALYSIS ON RANDOM FOREST

We randomly select 179, 169 and 111 files for test and the rest for training in EN1, EN2 and EN3 respectively. The relations between subjective measurement and predicted measurement are shown in Fig. 8. It can be seen that our approach can give a satisfactory performance under three conditions. We adopt cosine distance to represent the relations between real sub. and predicted sub., which is 0.9868, 0.9701, and 0.9941 separately. Since test data number and distribution among three conditions are different, shown in Fig. 6, the expressions of results are a little distinct. For example, compared with subfigure (a) and (b) in Fig. 8, the data points in subfigure (c) are concentrated in the upper right corner. The main reason lies in that the difficulty of collecting low level IDL data increases in outdoor real communication environment. From Fig. 6, it can be seen that unlike EN2, data on IDL 1 and IDL 2 are relatively sparse in EN3.

Fig. 9 draws the absolute predicted error under three environments. It gives the file number when the corresponding error under threshold 1 and 0.5 respectively. For EN1, total file number is 179, and 168 files (93.9%) under threshold 1, 134 files (74.9%) under threshold 0.5. While for EN2, the corresponding number is 169, 151 (89.3%) and 128 (75.7%). EN3 is outdoor real environment, and the test data only contains 111 files. Among those 111 files, there are 108 files with less error than threshold 1, occupying about 97.3% percent. Additionally, there are still 81.1% files satisfying the predicted error within 0.5.

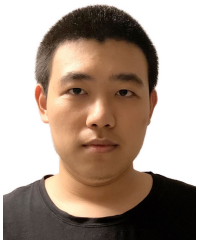
V. CONCLUSION

We presented a new open assessment benchmark for extreme communication environment with ultralow SNR. By a new subjective assessment named information damage level (IDL), which is IDL is kind of combination of objective and subjective assessment, we adopt random forest to bridge the gap between subjective and objective assessments. IDL limits the grade level by subjective recognition rate (SRR), and only gives tester the freedom to score within grade. It can avoid the wrong assessment across grades, which often happens with serious active interference. Experimental results show the effectiveness of the proposed method on our dataset, which provides a beneficial way to promote the innovation in speech or environment assessment.

REFERENCES

[1] H. C. Taal, C. R. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proc. ICASSP*, 2010, pp. 4214–4217.
[2] P. Janbakhshi, I. Kodrasi, and H. Bourlard, "Pathological speech intelligibility assessment based on the short-time objective intelligibility measure," in *Proc. ICASSP*, 2019, pp. 6405–6409.
[3] P. Janbakhshi, I. Kodrasi, and H. Bourlard, "Spectral subspace analysis for automatic assessment of pathological speech intelligibility," in *Proc. InterSpeech*, 2019, pp. 3038–3042.
[4] J. Dennis, H. D. Tran, and E. S. Chng, "Image feature representation of the subband power distribution for robust sound event classification," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 2, pp. 367–377, Feb. 2013.
[5] Y. Yuhong, Y. Hongjiang, H. Ruimin, G. Li, and X. Songbo, "Auditory attention based mobile audio quality assessment," in *Proc. ICASSP*, 2014, pp. 1389–1393.
[6] J. H. Flesner, D. S. Ewert, B. Kollmeier, and R. Huber, "Quality assessment of multi-channel audio processing schemes based on a binaural auditory model," in *Proc. ICASSP*, 2014, pp. 1340–1344.
[7] J.-H. Flesner, T. Biberger, and S. D. Ewert, "Subjective and objective assessment of monaural and binaural aspects of audio quality," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 7, pp. 1112–1125, Jul. 2019.
[8] G. Mittag and S. Moller, "Non-intrusive speech quality assessment for super-wideband speech communication networks," in *Proc. ICASSP*, 2019, pp. 7125–7129.
[9] J. Ooster, R. Huber, and T. B. Meyer, "Prediction of perceived speech quality using deep machine listening," in *Proc. Interspeech*, 2018, pp. 976–980.
[10] S. W. Fu, Y. Tsao, H.-T. Hwang, and H. Wang, "Quality-net: An end-to-end non-intrusive speech quality assessment model based on BLSTM," in *Proc. Interspeech*, 2018, pp. 1873–1877.
[11] L. Zhang, T. Xiao, J. Hao, and X. Xiang, "Regression forest for interference assessment in real ultra short-wave communication jamming system," in *Proc. WCICA*, 2016, pp. 1459–1462.

LEI ZHANG (Member, IEEE) received the Ph.D. degree in computer science from the Harbin Institute of Technology (HIT), China, in 2004. She was a Professor of computer science with the College of Information and Communication Engineering, HIT, from 2005 to 2017. She is currently a Professor with the Guangdong University of Petrochemical Technology. Her research interests include signal/image processing, computer vision, and machine learning.



XILE ZHAO is currently a Student of the College of Computer Science and Technology, Guangdong University of Petrochemical technology. His research interests include signal/image processing, computer vision, and machine learning.

XIN LI received the Ph.D. degree in computer science from the Dalian University of Technology, China. He is currently an Associate Professor of the College of Computer Science and Technology, Guangdong University of Petrochemical technology. His research interests include signal/image processing, computer vision, and machine learning.