*INTERSPEECH 2020*
October 25–29, 2020, Shanghai, China

# A Sound Engineering Approach to Near End Listening Enhancement

*Carol Chermaz, Simon King*

The Centre for Speech Technology Research, University of Edinburgh, United Kingdom
c.chermaz@sms.ed.ac.uk

## Abstract

We present the beta version of ASE (the Automatic Sound Engineer), a NELE (Near End Listening Enhancement) algorithm based on audio engineering knowledge. Generations of sound engineers have improved the intelligibility of speech against competing sounds and reverberation, while maintaining high sound quality and artistic integrity (e.g., audio track mixing in music and movies). We try to grasp the essential aspects of this expert knowledge and apply it to the more mundane context of speech playback in realistic noise. The algorithm described here was entered into the Hurricane Challenge 2.0, an evaluation of NELE algorithms. Results from those listening tests across three languages show the potential of our approach, which achieved improvements of over 7 dB EIC (Equivalent Intensity Change), corresponding to an absolute increase of 58% WAR (Word Accuracy Rate).

**Index Terms**: speech modifications, Near End Listening Enhancement, sound engineering

## 1. Introduction

Speech communication via electronic devices is ever-present in our daily lives. Besides traditional devices like TV and radio, new means of communication - like videoconferencing - are focusing our attention on the issue of speech intelligibility, particularly when this is delivered through loudspeakers. In this scenario, NELE (Near End Listening Enhancement) is a useful tool for facilitating communication. In contrast to speech enhancement, which recovers speech from a noisy mixture, NELE modifies the clean signal *before* it is played back, with a view to the subsequent communication channel disturbances that will affect it (e.g., noise and reverberation).

Different approaches to NELE have been pursued over the past two decades; a range of algorithms is described in [1]. A common strategy is to modify speech in ways that mimic the natural changes humans make when they speak in noise: e.g., articulating words more carefully, hence increasing the consonant/vowel ratio; boosting the energy in the higher part of the spectrum, hence making formants more pronounced [2]).

In our previous study on the efficacy of NELE techniques [3] we evaluated three state-of-the-art algorithms in simulated real-world environments. Based on the insights we gained from that study, we created the beta version of ASE (the Automatic Sound Engineer), the NELE algorithm we submitted to the Hurricane Challenge 2.0 [4].

### 1.1. Motivation for our approach

While the scientific community is engaged in a fine-grained analysis of the various factors that influence intelligibility, a more immediate and holistic approach is adopted by the entertainment industry. The objectives of NELE - delivering the human voice clearly against competing sounds and reverberation – has been the focus of sound engineers ever since there has been such a profession. Whilst making choices driven by artistic intent, engineers typically adopt strategies with a scientific basis. For example, DRC (Dynamic Range Compression) attenuates the parts of the speech where energy is abundant, while amplifying those that are too quiet to be audible. This process improves the consonant/vowel ratio, as most of the energy in speech signals is found in the vowels. As another example, a pre-emphasis filter is used to boost energy in that part of the spectrum most significant for speech intelligibility: F1 and F2 are typically found in the 300-2200 Hz range, and consonants even higher than that [5]; however, most of the energy in speech recordings typically resides around F0, i.e., 60-400 Hz.

Sound engineers can achieve high intelligibility: in music productions, lyrics are typically highly intelligible, notwithstanding the instruments in the background and often large amounts of deliberately-added reverberation. This is true even though intelligibility is usually secondary to a much more complicated matter: the pleasantness of the ensemble. In contrast, intelligibility is the main focus of NELE, and the quality of sound is often a secondary issue at best. Tang *et al.* [6] explain how algorithms that are very effective against noise may be less pleasant than unmodified speech when heard at high SNR (i.e., favourable low noise conditions) and that this might be because noise masks the processing artefacts such that, while listening in noise, listeners perceive as more "pleasant" the speech that is simply more intelligible.

Whilst the entertainment industry utilises knowledge obtained through scientific research, the reverse seldom happens. The process of building software solutions based on expert knowledge is known as *knowledge engineering* and can be applied to the sound engineering domain [7]. Practitioners in the music industry tend to converge to the same artistic choices [8]; there appears to be a *target frequency contour for a mix*, although it seems to vary through the years and be influenced by genre; engineers seem also to express similar preferences for DRC [9]. Automation in music production has been explored in [9] and [10]; a review of different methods can be found in [11]. Our algorithm is focused on the human voice, with the specific goal of maximising intelligibility in noise and reverberation. This falls much closer to the endeavours of the hearing aid and telephony industries, which are much less concerned with aesthetic matters than the music industry. Nevertheless, ASE was conceived as a meeting of all of these realities, hopefully drawing the best characteristics from each.

## 2. Algorithm Design

In order to investigate the modifications a sound engineer may apply in a NELE scenario, we followed an empirical approach similar to [7]: let an expert (in this case the first author, who has a background in the field) perform the task and, from an analysis of output, then design an automatic algorithm.



http://dx.doi.org/10.21437/Interspeech.2020-2748

### 2.1. Generation of reference stimuli

We used an established commercial DAW (Digital Audio Workstation) with state-of-the-art plug-ins to process the speech signals needed for this analysis. Recordings of the Matrix sentences in three languages (English, German and Spanish) [12], [13], were provided as material for the Hurricane Challenge 2.0. We concatenated a number of utterances from all languages and imported them into the DAW. Since the provided corpora featured male speakers only, we used additional material from female voices [14, 15, 16] as well as male speakers from other corpora [14, 17], in order to minimise potential biases due to speaker and recording characteristics. Participants in the challenge were advised that entries would be processed with noise and reverberation, and mock-up stimuli were provided as an illustration. The challenge includes three reverberation conditions: far, medium and near. We used the provided cafeteria noise plus reverberation (in varying amounts) created with a plug-in to simulate our best estimate of the three conditions.

To enhance intelligibility, we processed the material with a parametric equalizer, a 4-band compressor, then a limiter/maximiser (a type of broadband compression). We modified the parameters until all speech tokens were highly intelligible (as judged by the sound engineer), even at low SNRs combined with long reverberation times.

Based on the results in [1] and [3], we selected SSDRC [18] as a benchmark NELE algorithm, used it to process the same speech tokens, then imported these into the DAW for comparison. We aimed to achieve the same intelligibility with our processing, while avoiding the unpleasant artefacts.

Within the time constraints of the Hurricane Challenge 2.0, it was only possible to evaluate quality informally with non-expert listeners, in order to confirm that our choices were also acceptable to a general audience. Listeners were exposed to stimuli processed by ASE and by SSDRC, in noise and in quiet, and asked which they preferred in each condition. Once we achieved comparable intelligibility along with a preference for our approach, we exported the processed speech signals for analysis.

### 2.2. Analysis of reference stimuli

The expert-created signals were analysed in order to design an algorithm that emulated the expert's processing. The algorithm has similar stages to those used by the expert: multi-band compression, equalization and broadband compression.

We divided our reference signals in six bands: the number of channels and frequency ranges being inspired by modern mixing consoles used in sound engineering. We used FIR filters (Matlab `fir1`) with a large number of taps to give a steep roll-off minimising band overlap and phase rotation. We analysed the power in each band and, rather than using absolute values, we defined the third band as reference (0 dB) and calculated the difference in dB for the other bands. The power scheme we obtained from this analysis is used in ASE to equalize the signal. An example of spectral power distribution for plain and modified speech can be seen in Figure 1.

An informal analysis of the temporal envelope of our reference stimuli revealed a substantial degree of natural variation, in spite of DRC. For this reason, we decided to try and match those variations with a mild approach to compression, inspired by common practice in sound engineering and hearing aid technology. In order to pursue an artefact-free sound, we built an ad-hoc compressor, taking advantage of the non-causal processing allowed by the use case.

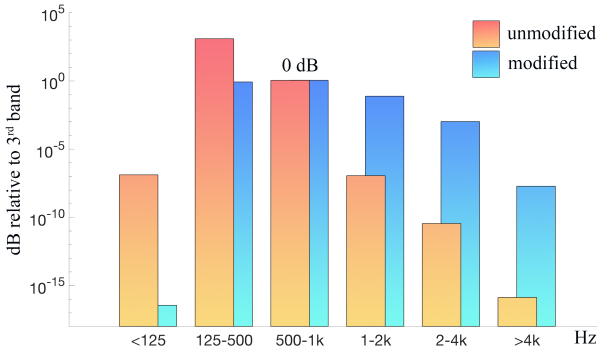


Figure 1: *Power distribution of a plain speech utterance (orange); the same utterance after processing with ASE beta (blue). Stimuli have the same RMS value.*

### 2.3. Non-causal dynamic range compression (DRC)

While compressors are widely used in industry, they are seldom explained in academic literature; a good summary of methods can be found in [19]. The typical parameters of DRC are threshold, ratio, attack and release time, knee width and make-up gain. Such parameters can be quite obscure for the lay user, but even experts have to choose on a case-by-case basis. One has to consider the specific characteristics of the signal at hand and the desired output; automatic choice of parameters is hence highly desirable. A model for automatic attack and release times was proposed in [20]; a method for compressing multiple tracks in a mix was proposed in [9].

As transients are difficult to tackle in real-time processing, compressors typically feature a side chain which contains the smoothing detector (which provides a *smoother* representation of the signal) and the gain computer; if using *look-ahead*, compression is calculated over a short time frame, and the output is hence delayed by this amount. There is a trade-off between compression quality and delay. In this work, we take advantage of a non-causal approach, since we are processing speech recordings; this equals unlimited *look-ahead*. For the Challenge, we processed individual utterances of five words.

In ASE beta, some compression parameters were fixed during the design process, and some are calculated on the input signal. Compression is instantaneous, based on a smoothed representation of the signal, which we call the *guide*. The guide can be *RMS* or *peak*, and we calculate it in the following way. The input signal is analysed in time-frames of 20 ms (every 10 ms); for each frame we calculate the RMS (for the *RMS guide*) or the maximum absolute value of the signal (for the *peak guide*). The obtained values are then interpolated with Matlab's `interp1q` function. An RMS guide can be seen in Figure 2. We chose an RMS guide for the lower frequency bands, since their temporal envelopes are slowly-varying, and peak for the higher bands (which vary more quickly).

In our compressor, the detector lies before the gain computer; the gain is hence based on the difference (in dB) between the guide and the threshold. The threshold is determined as a fraction the maximum value of the guide (in dbFS); see example in Figure 2. In the algorithm design process, we set those fractions as constants for each band; we allowed for a larger fraction (hence less compression) in the lower bands and reduced it gradually with increasing frequency. We fixed the compression ratio to 2 for all bands.

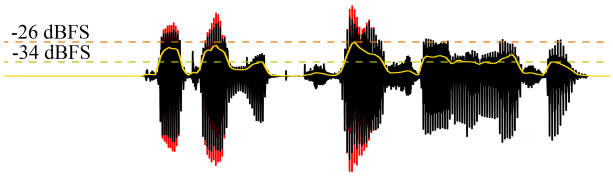


Figure 2: *DRC. Black: compressed signal; red: original signal; yellow: guide (RMS). The threshold (-34dBFS) is a fraction of the maximum value of the guide (-26 dbFS). Ratio = 2:1.*

### 2.4. Summary of processing steps in ASE beta

- Divide the signal into 6 frequency bands
- Compress the dynamic range of each band
- Scale bands according to power scheme
- Recombine signals, then broadband compression
- Rescale signal to original RMS

ASE beta is *noise-unaware* and *reverb-unaware*: stimuli are processed independently from channel conditions. This choice is motivated by the effectiveness and ease of use that SSDRC demonstrated in [3].

## 3. Subjective Evaluation

In the Hurricane Challenge 2.0, listening tests were run in German, English and Spanish, using recordings of Matrix sentences [12], [13]. 187 participants were recruited (German=62, English=63, Spanish=62). Speech stimuli were mixed with cafeteria noise in 3 reverberation conditions: near, medium and far. Each condition was evaluated at 3 SNRs: low, medium and high, expected to yield respectively 25%,50% and 75% WAR (Word Accuracy Rate) for plain (i.e., unmodified) speech. WAR measures intelligibility as the percentage of correct words listeners can recall after hearing a stimulus. For a detailed description of the experiments, please refer to [4].

Results are reported in terms of WAR in Figure 3. ASE showed a mean absolute improvement of +37% WAR across conditions. The benefit was larger in near reverberation, reaching +58% for Spanish at low SNR.

Although WAR gives an idea of the results at first glance, intelligibility gains are better expressed as EIC (Equivalent Intensity Change) [21], which is the amount in dB one would have to amplify unmodified speech in order to achieve the same intelligibility as modified speech. Algorithms behave differently at different SNRs, and for this reason one must assess their performance at different points along the psychometric curve [21, 3]. Since in the present experiment the reference values for plain speech are quite varied across conditions, instead of grouping results by SNR we decided to fit the data points to a psychometric curve (using Matlab `nlinfit` and the logistic function described in [22]) for every language/reverb condition, and then calculate the offset between the plain and ASE curves at the 50% point (hence the notation EIC 50). However, results must be taken with caution as the data points for ASE all lie at the top of the curve, which may lead to an incorrect estimate of its slope. This being said, ASE provided substantial benefit in all conditions, from an estimated minimum of 4.4 dB to a maximum of 7.3 dB (EIC 50), outperforming other entrants to [4] in 25 out of 27 conditions.

## 4. Discussion

### 4.1. Intelligibility

The difference in absolute WAR gains for ASE beta across different languages (in the same reverb/SNR condition) is mainly due to reference plain speech having different WAR scores. The EIC 50 differences across languages may be due to individual differences of the speakers [13] and recording procedure/equipment.

We assume the Hurricane Challenge 2.0 test environment to be comparable to [3], although we do not have exact knowledge of the impulse responses used in the Challenge. In [3], SSDRC achieved the best performance in cafeteria noise, with an EIC of 4.2 dB at a low SNR. In [1] the variant uwSSDRCt achieved an EIC of 5.1 dB in SSN (Speech Shaped Noise). In the present Challenge, ASE achieved 7.3 dB in cafeteria noise and reverberation, outperforming SSDRC in all conditions but one. This might be explained by the more fine-grained processing in ASE. While SSDRC aims at reducing spectral tilt via pre-emphasis filters, ASE aims at increasing the Speech Intelligibility Index [23] by scaling the power in separate frequency bands (see Figure 1). Additionally, ASE performs 2 stages of DRC, multi-band first and then broadband, while SSDRC only performs one stage of broadband DRC.

Results from [1], [3] and the present Challenge are not directly comparable. Even where noise conditions are similar, speech corpora and test methodologies are different. While [1] and [3] featured a recording of the Harvard Sentences in English [24, 17], the Hurricane Challenge 2.0 uses the Matrix sentences in three languages. Both Harvard and Matrix sentences have low semantic predictability, but while the former use a rather large vocabulary, the latter only use words from a closed set. While in [1] and [3] listeners typed what they heard into a free-response field, in the present Challenge listeners responded via a graphical interface only displaying words from the closed set. Results from a 2019 study (in preparation for publication) found that the psychometric curves for the two methods are comparable in slope (although having a significant offset). We therefore predict that ASE would achieve similar EIC gains in a free-response evaluation.

### 4.2. Quality of speech sound

In our informal survey, listeners' preference for our approach over the SSDRC benchmark is in part attributable to the frequency profile being controlled by equalization, which reduces the risk of excessive loudness in some frequency bands. Also, a moderate and carefully customized DRC guarantees absence of unpleasant artefacts. In this case, the non-causal approach (which for the Hurricane Challenge means look-ahead to the end of the current utterance) is an explicit advantage.

The effects of DRC parameters on intelligibility and quality are widely studied in the hearing prosthetics domain, but there seems to be no consensus on ideal values [25]. In ASE beta we set the compression ratio to 2 because low ratios ($\leqslant 2$) appear to be generally well tolerated [26]. The effect of ratio is intertwined with the threshold – but we were quite cautious here too, as listeners may opt for a higher threshold when given the choice [27].

Given the lack of agreement in hearing technology, we chose our number of channels and frequency ranges inspired by sound engineering hardware, where six bands is considered plenty - and in fact is available only on high-end mixing consoles. The idea proved to be successful in this case, although it

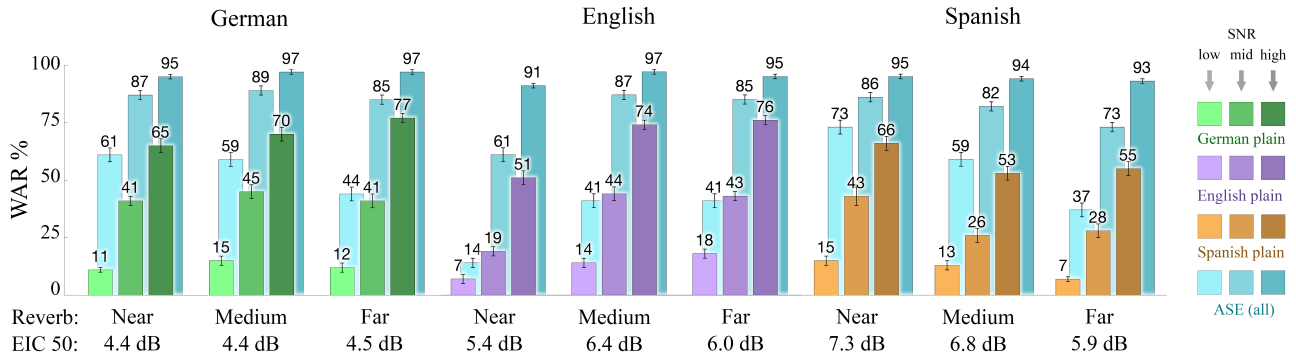


Figure 3: *WAR% (with standard error) for all conditions. Results are divided by language, reverberation (Near, Medium, Far) and SNR (low, mid, high). ASE scores for each condition are depicted in the adjacent blue columns. EIC 50 indicates the intelligibility gain provided by ASE. For a comparison with other algorithms, please refer to [4].*

would be interesting to try a different number of channels (and possibly different DRC setting).

The aim of ASE is to find a fine balance between intelligibility and quality, but defining *quality* is not trivial. Its intended use case is "high-fidelity sound", as in professional entertainment productions. Objective measures like PESQ [28] and POLQA[29] are designed to evaluate speech quality, but they require a *reference* signal: their objective is to assess the corruptions due to a communication channel (e.g., telephone line) by comparing the noisy output to the original input. ANIQUE [30] is a non-intrusive measure, but it is intended for narrowband signals only. Such measures are of little help in our use case, as we are looking to rate the quality of a hi-fidelity broadband signal as-is, in the absence of noise and without a reference signal.

In recent years there has been growing interest in using cognitive load as an objective measure of quality. It is possible that such a measure would be capable of revealing the effect of subtle changes in the signal, where a traditional subjective assessment of intelligibility or quality (like MOS) would fail [31]. From recent studies, it has also emerged that listeners do not necessarily prefer the frequency profiles that lead to the highest intelligibility scores [32]. The method proposed in [32] and the measurement of cognitive load are valuable tools for further development of ASE. "Perfect" quality can only be defined with respect to listeners.

### 4.3. Additional considerations

NELE is not intended as a substitute for the *mastering* process that is normally performed in the production of commercial audio, but to be complementary. In fact, while professionally-recorded speech may be enjoyed at its best with high quality equipment in an ideal environment, NELE algorithms can be used to make sure such speech remains intelligible in reverberant or noisy conditions too. Most NELE algorithms described in the literature to date are only intended for mundane tasks that do not require an artistic touch (e.g., processing public address announcements); however, we cannot exclude the use of ASE in artistic contexts.

We are also interested in the potential of ASE for improving the accessibility of speech for individuals with hearing loss. Other NELE algorithms have proven to be beneficial in this use case [33, 34]. According to the World Health Organization, there is a substantial mismatch in the numbers of hearing-impaired people that *could* benefit from medical assistance and those who actively *seek* it. Individuals are often deterred by the stigma of disability associated with a hearing device, and the high-cost of such devices [35]. While professional help is key to the well-being of a hearing-impaired person, more intelligible speech from playback devices would be a useful step forward in improving accessibility.

## 5. Conclusions

In this paper we present the beta version of ASE (the Automatic Sound Engineer), a NELE (Near End Listening Enhancement) algorithm based on sound engineering knowledge.

The algorithm was built following a *knowledge engineering* approach. We used established commercial sound engineering tools to process speech (from several corpora) in simulated noisy and reverberant conditions, with the goal of maximising intelligibility while preserving quality - intended as the *pleasantness* of sound in professional entertainment productions. We analysed the frequency and temporal profile of the obtained expert tokens and tried to reproduce it in ASE, by means of multiband and broadband DRC, equalization and limiting. We built an ad-hoc non-causal compressor with some degree of automation, which we programmed for a relatively mild compression.

ASE was entered into the Hurricane Challenge 2.0, a multilanguage comparison of NELE algorithms in realistic noise and reverberation. ASE outperformed competitors in 25/27 conditions, showing improvements of over 4 dB EIC (Equivalent Intensity Change) in every scenario, and up to 7.3 dB EIC. We find this approach very promising, and we are developing a fully automated version of our algorithm: those parameters we set as constants in the design process of ASE beta (e.g., the compression ratio), will be computed case-by-case on the input signal.

## 6. Acknowledgements

The authors thank the organizers of the Hurricane Challenge 2.0 for performing the evaluations and Andreas Volgenandt for useful discussions on DRC. This project has received funding from the EU's H2020 research and innovation programme under the MSCA GA 67532 (the ENRICH network: www.enrich-etn.eu).
Audio files are available at
`http://homepages.inf.ed.ac.uk/s1758351/ASEbeta.html`

# 7. References


[1] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Intelligibility-enhancing speech modifications: the Hurricane Challenge," in *Proc. Interspeech*, Lyon, France, August 2013.

[2] J. C. Junqua, S. Fincke, and K. Field, "The Lombard effect: a reflex to better communicate with others in noise," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, vol. 4, 1999, pp. 2083–2086 vol.4.

[3] C. Chermaz, C. Valentini-Botinhao, H. Schepker, and S. King, "Evaluating Near End Listening Enhancement Algorithms in Realistic Environments," in *Proc. Interspeech 2019*, 2019, pp. 1373–1377.

[4] J. Rennies, H. Schepker, C. Valentini-Botinhao, and M. Cooke, "Intelligibility-enhancing speech modifications – the hurricane challenge 2.0," in *Proc. Interspeech*, October 2020.

[5] P. C. Loizou, *Speech enhancement: theory and practice*. CRC press, 2013, ch. 3, pp. 59–65.

[6] Y. Tang, C. Arnold, and T. Cox, "A study on the relationship between the intelligibility and quality of algorithmically-modified speech for normal hearing listeners," *Journal of Otorhinolaryngology, Hearing and Balance Medicine*, vol. 1, no. 1, p. 5, 2018.

[7] P. Pestana and J. Reiss, "Intelligent audio production strategies informed by best practices," in *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*, Jan 2014.

[8] P. D. Pestana, Z. Ma, J. D. Reiss, A. Barbosa, and D. A. A. Black, "Spectral characteristics of popular commercial recordings 1950-2010," in *Audio Engineering Society Convention 135*, Oct 2013.

[9] Z. Ma, B. De Man, P. D. L. Pestana, D. A. A. Black, and J. D. Reiss, "Intelligent multitrack dynamic range compression," *J. Audio Eng. Soc*, vol. 63, no. 6, pp. 412–426, 2015.

[10] E. Perez-Gonzalez and J. Reiss, "Automatic equalization of multichannel audio using cross-adaptive methods," in *Audio Engineering Society Convention 127*, Oct 2009.

[11] B. De Man, J. D. Reiss, and R. Stables, "Ten years of automatic mixing," in *3rd Workshop on Intelligent Music Production*, September 2017.

[12] B. Kollmeier, A. Warzybok, S. Hochmuth, M. A. Zokoll, V. Uslar, T. Brand, and K. C. Wagener, "The multilingual matrix test: Principles, applications, and comparison across languages: A review," *International Journal of Audiology*, vol. 54, no. sup2, pp. 3–16, 2015.

[13] S. Hochmuth, B. Kollmeier, and B. Shinn-Cunningham, "The relation between acoustic-phonetic properties and speech intelligibility in noise across languages and talkers," in *Proc. DAGA, German Annual Conference on Acoustics*, Munich, Germany, 2018, pp. 628–629.

[14] J. Yamagishi, C. Veaux, and K. MacDonald, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit (version 0.92)," 2019. [Online]. Available: https://doi.org/10.7488/ds/2645

[15] S. King and V. Karaiskos, "The Blizzard Challenge 2011," in *The Blizzard Challenge Workshop*. The University of Edinburgh, 2011.

[16] C. Valentini-Botinhao and J. Yamagishi, "Alba speech corpus," 2019. [Online]. Available: https://doi.org/10.7488/ds/2506

[17] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Hurricane natural speech corpus," 2013. [Online]. Available: https://doi.org/10.7488/ds/2482

[18] T. C. Zorilă, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in *Proc. Interspeech*, Portland, USA, September 2012.

[19] D. Giannoulis, M. Massberg, and J. D. Reiss, "Digital dynamic range compressor design—a tutorial and analysis," *Journal of the Audio Engineering Society*, vol. 60, no. 6, pp. 399–408, 2012.

[20] ——, "Parameter automation in a dynamic range compressor," *Journal of the Audio Engineering Society*, vol. 61, no. 10, pp. 716–726, 2013.

[21] M. Cooke, C. Mayo, C. Valentini-Botinhao, Y. Stylianou, B. Sauert, and Y. Tang, "Evaluating the intelligibility benefit of speech modifications in known noise conditions," *Speech Communication*, vol. 55, no. 4, pp. 572–585, 2013.

[22] T. Brand and B. Kollmeier, "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," *The Journal of the Acoustical Society of America*, vol. 111, no. 6, pp. 2801–2810, 2002.

[23] "Methods for calculation of the speech intelligibility index," American National Standard, New York, USA, Standard, 1997.

[24] IEEE, "IEEE recommended practice for speech quality measurement," *IEEE Trans. on Audio and Electroacoustics*, vol. 17, no. 3, pp. 225 – 246, 1969.

[25] J. M. Kates, "Understanding compression: Modeling the effects of dynamic-range compression in hearing aids," *International Journal of Audiology*, vol. 49, no. 6, pp. 395–409, 2010.

[26] A. C. Neuman, M. H. Bakke, S. Hellman, and H. Levitt, "Effect of compression ratio in a slow-acting compression hearing aid: Paired-comparison judgments of quality," *The Journal of the Acoustical Society of America*, vol. 96, no. 3, pp. 1471–1478, 1994.

[27] C. Barker and H. Dillon, "Client preferences for compression threshold in single-channel wide dynamic range compression hearing aids," *Ear and Hearing*, vol. 20, no. 2, pp. 127–139, 1999.

[28] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, vol. 2. IEEE, 2001, pp. 749–752.

[29] J. G. Beerends, C. Schmidmer, J. Berger, M. Obermann, R. Ullmann, J. Pomy, and M. Keyhl, "Perceptual objective listening quality assessment (polqa), the third generation itu-t standard for end-to-end speech quality measurement part i—temporal alignment," *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 366–384, 2013.

[30] D.-S. Kim, "Anique: an auditory model for single-ended speech quality estimation," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 821–831, 2005.

[31] A. Govender and S. King, "Using pupillometry to measure the cognitive load of synthetic speech," in *Proc. Interspeech 2018*, 2018, pp. 2838–2842.

[32] O. Simantiraki, M. Cooke, and Y. Pantazis, "Effects of spectral tilt on listeners' preferences and intelligibility," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 6254–6258.

[33] J. Rennies, J. Drefs, D. Hülsmeier, H. Schepker, and S. Doclo, "Extension and evaluation of a near-end listening enhancement algorithm for listeners with normal and impaired hearing," *The Journal of the Acoustical Society of America*, vol. 141, no. 4, pp. 2526–2537, 2017.

[34] T.-C. Zorilă, Y. Stylianou, S. Flanagan, and B. C. J. Moore, "Evaluation of near-end speech enhancement under equal-loudness constraint for listeners with normal-hearing and mild-to-moderate hearing loss," *The Journal of the Acoustical Society of America*, vol. 141, no. 1, pp. 189–196, 2017.

[35] B. Shield, "Evaluation of the social and economic costs of hearing impairment. a report for hear-it," 2006. [Online]. Available: https://www.hear-it.org/sites/default/files/multimedia/documents/Hear_It_Report_October_2006.pdf