4th International Workshop on Perceptual Quality of Systems (PQS 2013)
2-4 September 2013, Vienna, Austria

A Neurogram Matching Similarity Index (NMSI) for the Assessment of Audio Quality

Michael Drews[1], *Steffen Schlapak*[1], *Stefano Rini*[2], *Michele Nicoletti*[1], *Werner Hemmert* [1]

[1]Institute for Medical Engineering, Technische Universität München, Munich, Germany
E-mail: {michael.drews, steffen.schlapak, michele.nicoletti, werner.hemmert}@tum.de
[2] Department of Electrical Engineering, Stanford University, Stanford, CA, USA
E-mail: stefano@wsl.stanford.edu

Abstract

In this paper, the performance of neurogram and spectrogram similarity indexes for the prediction of perceived audio quality is evaluated. Additionally, a new method for comparison of internal signal representations is introduced that relies on a two-dimensional extension of the Needleman-Wunsch algorithm. It approximates the two-dimensional edit distance which is given by the minimum cost to transform one matrix into another by sequentially inserting, deleting and changing the positions and values of single elements. By choosing the cost of each operation, we define a similarity index which can be used for assessment of audio quality among neurograms or spectrograms. We evaluate the performance of this measure to estimate the sound quality of audio files degraded by low bit-rate audio COder-DECoders (CODECs). The new measure shows high correlation with Perceptual Evaluation of Audio Quality (PEAQ) predictions and outperforms other measures of similarity in the literature. We find similarity of spectrograms and neurograms to be sensitive to changes in audio quality for low bit-rate codings.

Index Terms: audio quality, edit distance, low bit-rates

1. Introduction

Objective measures of audio quality try to detect and quantify the audibility of artifacts by rating the difference between a degraded test signal and an undegraded reference signal in a way that is relevant to the subjective quality estimations of a human listener. For most approaches, the predictions are based upon the comparison of so-called "internal representations", usually two-dimensional spectro-temporal signal representations that reflect psychoacoustically relevant effects such as frequency masking or thresholding. From a physiological point of view, an audio signal that enters the ear is processed by the sound conducting mechanisms of the auditory periphery and ultimately transformed into neuronal discharge patterns of the auditory nerve fibers (ANFs). All signal information which is not encoded in the discharge characteristics of the auditory nerve, is hence lost for further processing in higher cognitive stages. Under the assumption that many of the perceptual effects which are involved in hearing can be explained by the sound processing in the auditory periphery [1], it should therefore be possible to employ neural signal representations for the prediction of perceptual values such as audio quality. There have been several approaches to the related problem of predicting speech intelligibility from neural signals. The Neural Articulation Index (NAI) [2] rates signal differences by the cross-correlation of time dependent mean discharge rate functions of single ANFs. The Neurogram Similarity Index Measure (NSIM) [3] was used to predict speech intelligibility on basis of an image processing approach for comparison of neurograms. Neurograms contain the neural response of the inner ear over the whole tonotopic map and can be obtained by physiologically inspired computational models of the auditory periphery [4, 5]. The NSIM compares neurograms by evaluating luminance, contrast and structure differences between two neurograms which are considered to be images. Recently, the NSIM has been applied within the Virtual Speech Quality Objective Listener (ViSQOL) [6] framework to predict audio quality based on spectrograms. Other methods for prediction of audio quality do not necessarily base upon physiologically accurate simulations of the auditory periphery. In PErception MOdel based Quality estimation (PEMO-Q) [7] an "effective" model of auditory periphery [8] is applied and signal differences are evaluated by the cross-correlation of the filter outputs from an analysis modulation filter bank. The most common method for audio quality prediction, PEAQ [9], combines several similarity measures (11 in the basic version) such as envelope modulation and noise-to-masker ratio to produce a single similarity measure by training a neural network with subjective judgements. Internal signal representations are generated here with a Fast Fourier Transform (FFT) based ear model. In the advanced version of this tool, additionally a filter

Supported by within the Munich Bernstein Center for Computational Neuroscience by the German Federal Ministry of Education and Research (reference number 01GQ1004B).

10.21437/PQS.2013-19

bank model is used for the ear model, together with more involved similarity measures in the training of the neural network. The PEAQ achieves high correlation scores with subjective quality ratings by combining the benefits from several feature extraction techniques into one single measure. In this paper, we introduce the Neurogram Matching Similarity Index (NMSI) [10] as a new method for assessment of similarity among two-dimensional internal signal representations. The NMSI estimates the edit distance between matrixes which is defined by the minimum number of operations to transform one matrix into another by sequentially applying element insertions, deletions or substitutions. The edit distance can be assessed via an extension of the Needleman-Wunsch algorithm [11] for one-dimensional sequence comparison. NMSI is applied to neurograms and to spectrograms and the performance of the respective method for prediction of audio quality is evaluated in terms of correlation with objective quality scores predicted with the PEAQ. Results are compared with predictions from other methods for neurogram and spectrogram similarity in the literature.

Contributions: We introduce the NMSI as a measure for similarity of neurograms and spectrograms. We perform an optimization of the parameters of this measure and show that this measure can be used for the task of predicting audio quality of audio files degraded by low bit-rate audio CODECs. The performance of this and other methods from the literature is evaluated using neurograms as well as spectrograms as internal signal representations.

2. Background

2.1. Neurograms

Neurograms were generated using the inner ear model by Holmberg and Hemmert [5], which was designed to mimic the human performance as close as possible. This model generates neuronal spike trains of single ANFs at any characteristic frequency (CF) of the inner ear. It includes a hydromechanical model of basilar membrane motion, outer hair cell amplification and a biophysical model of auditory nerve synapses which is able to reproduce the characteristics of high spontaneous rate (HSR), medium spontaneous rate (MSR) and low spontaneous rate (LSR) fibers. Spike trains are produced with a time resolution of 10μs and for 100 different CFs between 50 Hz and 14 kHz. We consider a composition of 60 % for HSR, 25 % for MSR and 15 % for LSR fibers in accordance with physiological data (from cat [12]).

Neurograms were downsampled to 23 frequency channels and 10 ms time bins by using square cosine windows. A time-frequency patch in a neurogram then displays the mean discharge rate at the center frequency of the corresponding channel at a time. Hence, we effectively obtain a coarse *rate-place-code* of neural coding ([13], [14]) which comprises the information of the signal's envelope. However, note that the NMSI is also applicable to neurograms with higher time and frequency resolution. Downsampling is conducted for a sufficiently smooth representation and to reduce aliasing and computational complexity.

2.2. Spectrograms

Spectrograms were generated using a Short-Time Fourier Transform (STFT) in 20 ms Hamming windows with 50% overlap. To obtain a spectrogram representation similar to that of a pre-processed neurogram, the STFT magnitude was binned in 23 logarithmically distributed and scaled frequency bins. Spectrograms were then linearly scaled to the dynamic range of the neurograms so that the NMSI parameters can be expected to have the same order of magnitude for both neurograms and spectrograms.

2.3. Data Collection

The audio samples used in our experiments are from the Sound Quality Assessment Material[1] (SQAM) recordings of the European Broadcasting Union, that comprises 70 audio files of solo instruments as well as more complex music and speech samples. Two seconds were selected from each file for further processing. A few files were omitted from the dataset because they were too short or too simple to be reduced in quality significantly. The remaining 59 files were encoded in MPEG-2 Audio Layer III (MP3) using the LAME Codec v3.99.5 with 14 different constant bit-rates between 8 kbps and 160 kbps. The samples were low-pass filtered at 14 kHz in order to match the frequency mapping of the auditory model. The files were then decoded to wav-format, some files had to be resampled to 48 kHz sampling frequency. Reference samples were generated by only low-pass filtering the original files at 14 kHz. Finally, a possible time-delay between the reference signal and the distorted one was eliminated by shifting the distorted signal by the time lag which was calculated using the cross correlation function. All files were set to a level of 65 $\mathrm{dB_{SPL}}$ (SPL: sound pressure level, relative to 20 μPa).

2.4. Audio Quality

Audio quality is typically assessed in listener tests performed in accordance with ITU-R Recommendation BS.1116. The perceived quality of a test signal is rated in Subjective Difference Grades (SDGs) relative to a reference signal on a continuous five-point scale ranging from -4.0 (very annoying) to 0.0 (imperceptible). Instead of SDGs, which were not available for this dataset, Objective Difference Grades (ODGs) from the PEAQ (basic version) were used in this paper for the optimization and tests of the proposed methods of audio quality prediction.

[1] http://tech.ebu.ch/publications/sqamcd

3. Neurogram Matching Transformation Index

3.1. The Two Dimensional Levenshtein (2DL) Algorithm

The Needleman-Wunsch algorithm [11] - also known as Sellers algorithm [15] - computes the edit distance between one-dimensional sequences and determines the optimal sequence of single-element-operations to minimize the cost of changing one sequence to another. There are three possible operations: insertion, deletion and substitution and a cost is assigned to each of them. To adapt the algorithm to our application, we use a simple variation in which the cost of substituting an element depends linearly on the absolute value of the difference between the elements. This approach has already been successfully applied to one-dimensional spike time sequences of single neurons [16]. We focus on a two dimensional extension of the Needleman-Wunsch algorithm, the 2DL algorithm, which was first proposed independently by Moore [17] and by Tanaka and Kikuchi [18]. For this framework we introduce the three cost parameters $\mathrm{q_t}$ and $\mathrm{q_f}$ for shifting a spectro-temporal element in a neurogram in time or frequency and $\mathrm{q_a}$ for changing its amplitude. The choice of these cost parameters controls the sensitivity of the measure to distortion in the temporal or spectral domain and to changes in the intensity of a neurogram (see Figure 1). By carefully choosing the cost of each transformation, it is possible to estimate experimental sound quality evaluation from similarity among two-dimensional internal signal representations. We define such estimator the Neurogram Matching Similarity Index (NMSI).

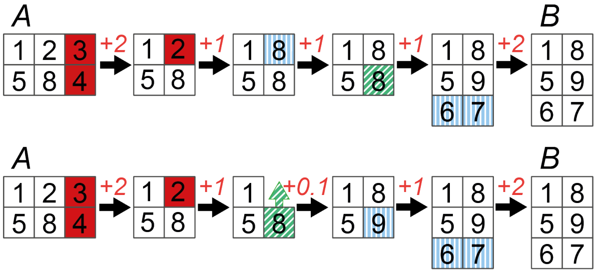

Figure 1: Example of the 2DL algorithm transforming one matrix into another. Elements highlighted with dark areas are deleted, diagonally hatched ones are substituted and vertically hatched ones are inserted. *Top:* Cost parameters $(\mathrm{q_t}, \mathrm{q_f}, \mathrm{q_a}) = (1, 1, 1)$. Total transformation cost is equal to 7. *Bottom:* Cost parameters $(\mathrm{q_t}, \mathrm{q_f}, \mathrm{q_a}) = (1, 0.1, 1)$. Total cost is 6.1 because reducing the cost of a vertical (frequency) shift $\mathrm{q_f}$ makes it more favourable to vertically shift the "8" up instead of deleting and inserting it.

3.2. Predicting Audio Quality

The NMSI scores were evaluated using both neurograms and spectrograms as internal signal representations, thus constituting two different approaches which we will formally denote as neuroNMSI and spectNMSI. The NMSI and PEAQ scores were evaluated for all test samples and bitrates which constitutes a set of about 800 different sample points. Part of the data was used for the fitting of a transfer function that maps NMSI values to the ODG scale of the PEAQ (see figures 4 and 5). This training data was chosen so as to contain all the different kinds of audio samples: music, speech as well as single instruments. The NMSI values of the remaining data were used for estimation and employed the reference function from above. The quality of the estimation between the predicted PEAQ-ODGs using the NMSI and the PEAQ is inferred by the correlation coefficient ρ between the predicted and the actual PEAQ.

The prediction of the NMSI is compared with the one provided by the NSIM [3]. To this end, the previous steps were repeated for the NSIM for both neurograms and spectrograms, constituting two new similarity metrics: neuroNSIM and spectNSIM. NSIM values are scaled between $\mathrm{d_{NMSI}} = 0$ (highest dissimilarity) and $\mathrm{d_{NMSI}} = 1$ (highest similarity). In this analysis NSIM scores were first re-mapped with the function

$$\mathrm{NSIM} = -\ln|\mathrm{d_{NMSI}} - 1|$$

to re-compensate the exponential growth of pure NSIM scores with higher bit-rates, which results in a higher degree of similarity among neurograms. Furthermore, we defined a trivial metric, ABS, by taking the sum over the absolute values of the difference between reference and test neurogram or spectrogram and logarithmizing the result.

$$\mathrm{ABS} = \ln\left(\sum_{\mathrm{i}} |(\mathrm{REF_i} - \mathrm{TEST_i}|\right)$$

This constitutes two additional similarity metrics, neuroABS and spectABS.

4. Results

4.1. Parameter Tuning

NMSI predictions depend on the choice of the three cost values $\mathrm{q_t}$, $\mathrm{q_c}$ and $\mathrm{q_a}$. Therefore, a parameter optimization was conducted first in order to obtain maximum correlation of NMSI scores with PEAQ predictions of audio quality. The cost of the insertion and deletion operations of single elements were set to $\mathrm{c_{insert}} = 1$ and $\mathrm{c_{delete}} = 1$ while the other cost of a substitution was assumed to be smaller than two. This cost constraint is intuitively clear: if this were not the case a substitution operation could be replaced by a deletion followed by an insertion. This will result in a metric that counts differences in the number of elements among realizations. Consequently, the cost parameters $\mathrm{q_t}$, $\mathrm{q_f}$ and $\mathrm{q_a}$ were also set to be smaller or equal to two.

In the following simulations, it was assumed that $q_t = 2$, which effectively prohibits any time-shift operations. In accordance with [10] this can be justified by the coarse time resolution of the neurograms and spectrograms (10 ms) and by the cross-correlation step in data pre-processing which minimizes any temporal misalignments between reference and test samples. For optimization of the remaining parameters q_f and q_a the penalty function was set to $1 - \rho_t$, with the correlation coefficient

$$\rho_t(y, f(x)) = \frac{C(f(x), y)}{\sqrt{C(f(x), f(x))C(y, y)}}$$

where C is the covariance matrix between the vector y, containing the PEAQ-ODGs, and the vector f(x), containing the NMSI values of the training data, re-mapped with the corresponding transfer function f. The position of the maximum was numerically determined using *Powell's method* [19]. Powell's method is an algorithm for finding the local minimum of a continuous function without calculating derivatives and thus suited for computationally intensive problems like the one at hand. It iteratively computes the minimum of the function along a given set of direction vectors. In one iteration of the algorithm, the function is subsequently minimized along all direction vectors and the last vector is substituted by the mean displacement vector of all minimizations performed in that iteration. Figures 2 and 3 show the correlation as a function of the cost q_t and q_a for two different parameter regimes (for neurograms and spectrograms). The figure also plots (black dotted line) the numerical optimal solution as across multiple iterations of the algorithm. Both figures indeed show that the objective function is smooth and the convergence to a numerical minimum sufficiently quick. Optimal prediction of PEAQ-ODGs was obtained with a parameter set of $(q_t, q_c, q_a) = (2, 0.6215, 0.1582)$ for neurograms and $(q_t, q_f, q_a) = (2, 0.5064, 1.5824)$ for spectrograms.

4.2. Correlation analysis

In total, we evaluated audio quality predictions based on six different similarity measures, namely spectNMSI and neuroNMSI, spectNSIM and neuroNSIM and spectABS and neuroABS. The transfer functions for these metrics are shown for neurograms in figure 4 and for spectrograms in figure 5. Looking at the transfer functions of spectNMSI and neuroNMSI we notice a flat part at low PEAQ scores which corresponds to low bit-rates. From this figure we notice that the NMSI grows here faster than the PEAQ which indicates a finer discrimination between low bit-rates than the PEAQ has here. The NSIM and the ABS show that behavior as well for low bit-rates but also for high bit-rates, in contrast to the NMSI which shows a linear dependence on PEAQ-ODGs here. This is partly due to the logarithm that was used for better representation of NSIM and ABS scores, whose transfer functions have a similar curve shape therefore. Table 1 lists the correlation coefficients obtained by applying the transfer functions to the validation data. The spectNSIM has the lowest correlation with the PEAQ for the validation data, although the correlation coefficient is comparable with the other methods for the training data. For neurograms all metrics yield comparable results. Using spectrograms as signal representations, we observe a higher correlation for the NMSI than for the NSIM and the ABS. The spectNMSI correlates best with PEAQ scores for both training and validation data.

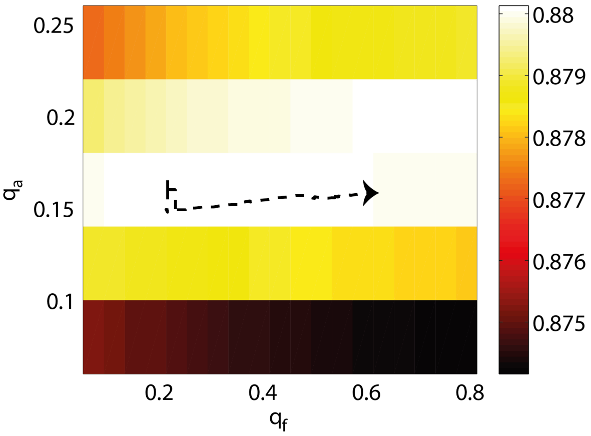

Figure 2: Correlation coefficient for neuroNMSI for a small parameter range. Black dotted line shows the pathway which is taken by the optimization algorithm.

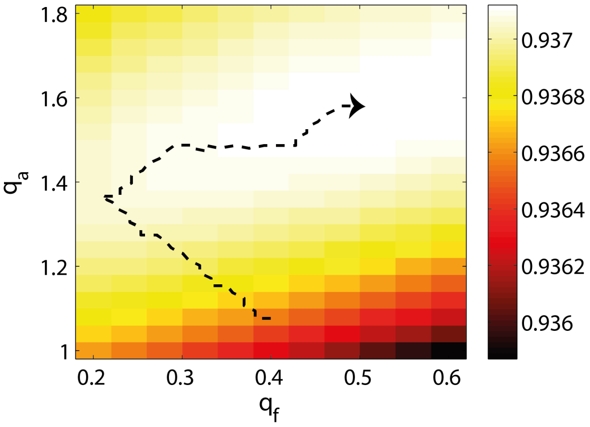

Figure 3: Correlation coefficient for spectNMSI for a small parameter range. Black dotted line shows the pathway which is taken by the optimization algorithm.

ρ_v	*neuro*	*spect*
NMSI	0.8985	0.9207
NSIM	0.8901	0.8520
ABS	0.8932	0.8981

Table 1: Correlation coefficients for the validation data.

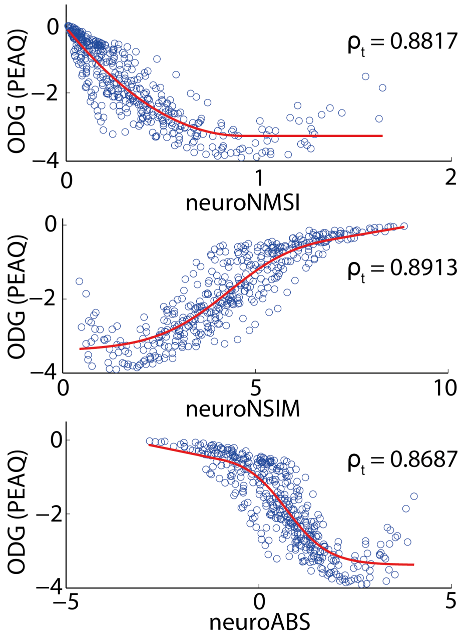

Figure 4: Transfer functions for neurograms and correlation coefficient ρ_{t} for the training data. *Top:* spectNMSI. *Middle:* spectNSIM. *Bottom:* spectABS.

5. Discussion

In this investigation, we used the correlation coefficient, which relies on the assumption of a linear dependence between two variables, as a measure of similarity between two different scales. We chose this measure because of ease in evaluation, however it is important to notice that this approach only reveals linear dependencies. Indeed, other possible approaches could reveal finer, non-linear dependencies between the variables.

Figure 6 shows the bit-rate dependent mean and variance of NMSI scores for cost parameters for two optimized parameter sets: $(\mathrm{q_t}, \mathrm{q_c}, \mathrm{q_a}) = (2, 0.6215, 0.1582)$ (dashed lines, optimized for neurograms) and $(\mathrm{q_t}, \mathrm{q_f}, \mathrm{q_a}) = (2, 0.5064, 1.5824)$ (solid lines, optimized for spectrograms). Curves can be compared only for the same parameters. For a given parameter setting, the difference between the neurogram (gray) and the spectrogram (black) curve is approximately given by the variance of the neuroNMSI curve for high and medium bit-rates. Considering the stochastic nature of neurograms, we can explain this relation as follows. A neurogram displays the mean discharge rate of the frequency selective ANFs, which is correlated with the spectral energy distribution of the signal given by a spectrogram. The stochastic nature of neural signals though adds a random process on top of this intensity map. Hence, the deviation between two neurograms can be approximated by the deviation between two spectrograms plus an additional deviation due to the stochasticity of neurograms. Under these assumptions, the distance between neurograms should be given by the distance between spectrograms plus the variance of the random process. In figure 6, this holds for high and medium bit-rates greater than approximately 32 kbps. With lower bit-rates though, the spectNMSI grows faster than the neuroNMSI which leads to the conclusion that neurogram distance here is mostly determined by the error of the MP3 coding of the signals than by stochastic fluctuations. The signal distortions induced by low-bit-rate codings are more visible in the spectrogram than in the neurograms due to the nature of MP3 codecs. We therefore predict the NMSI to be a good quality estimator for low-bitrate compressed audio. As presented in figure 7, the PEAQ estimation of audio quality saturates at a constant ODG score for bit-rates below 32 kbps which is not intuitive since - based on personal judgements - there are still big quality differences between the differ-

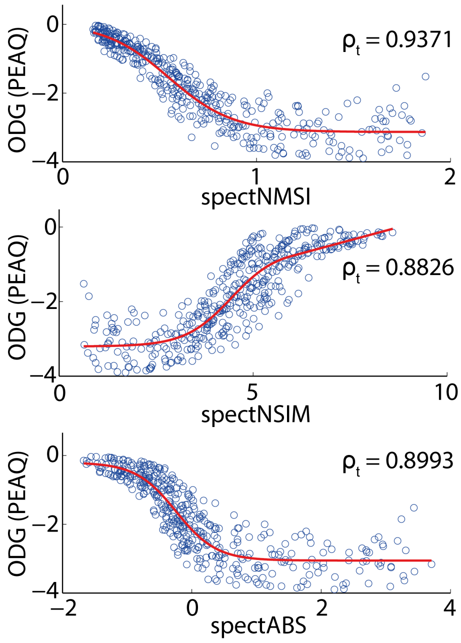

Figure 5: Transfer functions for spectrograms and correlation coefficient ρ_{t} for the training data. *Top:* spectNMSI. *Middle:* spectNSIM. *Bottom:* spectABS.

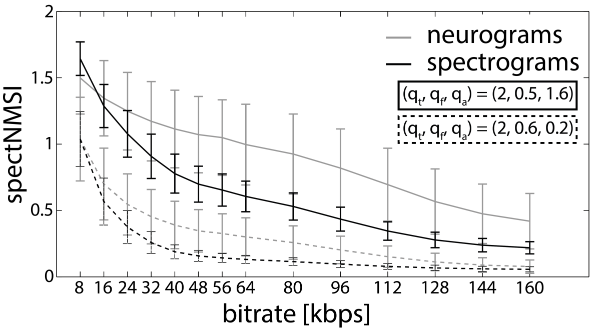

Figure 6: NMSI evaluated for various bit-rates and across samples.

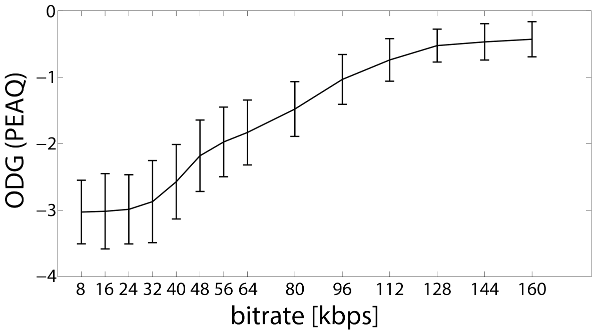

Figure 7: PEAQ evaluated for various bit-rates and across samples.

ent bit-rates. Eventual shortcomings of the PEAQ in this region have been described in [20] where the authors proposed an energy-equalization approach to overcome these deficits. We propose that the NMSI can be applied in this regime for prediction of audio quality and further investigations on this topic should be based on subjective quality ratings to enable direct comparison with PEAQ performance. The neuroNMSI approach is also appropriate for high quality audio when the stochasticity of the spike generation is reduced by longer or repeated signal representations.

The NSIM shows good correlation with PEAQ scores for both neurograms and spectrograms although it has no benefit over the trivial metric which detects only intensity differences in corresponding frequency-time patches. For the NSIM and the ABS there is no difference in performance between spectrograms and neurograms which also indicates that these metrics derive quality scores mainly on the base of differences in the spectral intensity between corresponding elements.

6. Conclusion

In this correspondence we introduced the Neurogram Matching Similarity Index (NMSI) as a new method for assessment of similarity among neurograms or spectrograms. We considered three different similarity indexes which are applied to both neurograms and spectrograms and showed that these indexes are highly correlated with other perceptual measures for estimation of audio quality from the literature. The underlying algorithms of the NMSI are not restricted to any particular assumptions about the time or frequency sampling of the neurograms and spectrograms. Hence, we see a high potential of the NMSI to evaluate also temporal fine structure cues contained in neurograms. This feature has the potential to extend application beyond predictions of perceived audio quality. Ongoing research involves subjective audio quality ratings from human listeners to further refine and validate the ability of the NMSI to capture audio quality and speech intelligibility.

7. References

[1] H. Fastl and E. Zwicker, Psychoacoustics: Facts and models, vol. 22, *Springer-Verlag New York Incorporated*, 2007.

[2] J. Bondy, I.C. Bruce, S. Becker, and S. Haykin, "Predicting speech intelligibility from a population of neurons," *Advances in Neural Information Processing Systems*, vol. 16, 2003.

[3] A. Hines and N. Harte, "Speech intelligibility prediction using a neurogram similarity index measure," *Speech Communication*, vol. 54, no. 2, pp. 306 – 320, 2012.

[4] M.S.A. Zilany, I.C. Bruce, P.C. Nelson, and L.H. Carney, "A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics," *The Journal of the Acoustical Society of America*, vol. 126, pp. 2390, 2009.

[5] M. Holmberg and W. Hemmert, "An auditory model for codingspeech into nerve-action potentials.," in *Joint Congress CFA/DAGA04 (2004)*, 2004, pp. 773–775.

[6] A. Hines, J. Skoglund, A. Kokaram, and N. Harte, "Robustness of speech quality metrics to background noise and network degradations: Comparing ViSQOL, PESQ and POLQA," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

[7] R. Huber and B. Kollmeier, "PEMO-Q a new method for objective audio quality assessment using a model of auditory perception," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 6, pp. 1902–1911, 2006.

[8] R. Dau, D. Püschel, and A. Kohlrausch, "A quantitative model of the effectivesignal processing in the auditory system. i. model structure," *The Journal of the Acoustical Society of America*, vol. 99, pp. 3615, 1996.

[9] T. Thiede, W.C. Treurniet, R. Bitto, C. Schmidmer, R. Sporer, J.G. Beerends, and C. Colomes, "PEAQ - the ITU standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, 2000.

[10] M. Drews, S. Rini, M. Nicoletti, and W. Hemmert, "The neurogram matching similarity index (NMSI) for the assessment of similarites among neurograms," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

[11] S.B. Needleman, C.D. Wunsch, et al., "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443–453, 1970.

[12] M.C. Liberman, "Physiology of cochlear efferent and afferent neurons: direct comparisons in the same animal," *Hearing Research*, vol. 34, no. 2, pp. 179–191, 1988.

[13] M.B. Sachs and E.D. Young, "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," *The Journal of the Acoustical Society of America*, vol. 66, no. 2, pp. 470–479, 1979.

[14] E.D. Young, "Neural representation of spectral and temporal information in speech," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1493, pp. 923–945, 2008.

[15] P.H. Sellers, "On the theory and computation of evolutionary distances," *SIAM Journal on Applied Mathematics*, vol. 26, no. 4, pp. pp. 787–793, 1974.

[16] J.D. Victor and K.P. Purpura, "Nature and precision of temporal coding in visual cortex: a metric-space analysis.," *Journal of Neurophysiolgy*, vol. 76, no. 2, pp. 1310–1326, Aug 1996.

[17] R.K. Moore, "A dynamic programming algorithm for the distance between two finite areas," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, , no. 1, pp. 86–88, 1979.

[18] E. Tanaka and Y. Kikuchi, "A metric between pictures," *Trans. IEICE*, vol. 63, pp. 1018–1025, 1980.

[19] M.J.D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Computer Journal*, vol. 7, no. 2, pp. 155–162, 1964.

[20] C.D. Creusere, "Quantifying perceptual distortion in scalably compressed mpeg audio," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2003, vol. 1, pp. 265–269 Vol.1.