Physiologically Based Predictors of Speech Intelligibility

Ian C. Bruce
Postal:
Department of Electrical and Computer Engineering
McMaster University
1280 Main Street West
Room ITB-A213
Hamilton, Ontario L8S 4K1
Canada

Email:
ibruce@ieee.org

Speech intelligibility predictors are powerful tools for evaluating how a listening environment, signal processing, and hearing impairment affect speech communication.

Introduction

Just a century ago, researchers at AT&T's Western Electric Research Labs (later renamed Bell Labs) began a comprehensive research program to develop an objective predictor of speech intelligibility that would provide a tool for efficiently assessing different speech telecommunications systems (Allen, 1996). This work was largely directed by the first president of the Acoustical Society of America (ASA), Harvey Fletcher, who was later made an Honorary Fellow of the ASA and then awarded the Gold Medal from the Society. The metric they developed was termed the articulation index (AI), which fundamentally measured the level of speech that is received above any background noise in a set of independent frequency bands, that is, a signal-to noise ratio (SNR) or the threshold of audibility if there is no noise in a band (French and Steinberg, 1947; Fletcher and Galt, 1950). Some nonlinear properties of the human auditory periphery were included in an ad hoc fashion. The AI was developed based on an extensive set of perceptual experiments using nonsense syllables as well as the fundamental knowledge about human psychoacoustics and the physiology of the ear at that time period.

Although the AI was developed to evaluate speech communication systems such as the analog telephone network, it was soon seen to be a valuable tool for the fields of speech and hearing research and audiology. However, a proliferation of different simplifications of the original AI was developed for these diverse purposes, including a 1969 American National Standards Institute (ANSI) standard (Hornsby, 2004), such that it was difficult to compare results across different studies. This prompted the development of the speech intelligibility index (SII) as a standard that captured the main principles of the AI but allowed for certain variations in the calculation method and application (ANSI, 1997), including allowing for degradations due to hearing loss, higher than normal sound presentation levels, and the upward spread of masking. The SII has also been extended to deal with cases of fluctuating background noise (Rhebergen et al., 2006).

A limitation of the AI and SII is that there are a number of distortions to speech, such as peak clipping or reverberation, that are known to affect speech intelligibility but that cannot be formulated simply in terms of an SNR, and the original AI only considered certain distortions relevant to telephony. This led to the development of the speech transmission index (STI), which uses the AI framework but substitutes a measure of acoustic modulation transfer for the SNR (Steeneken and Houtgast, 1980). This is based on the premise that speech information is primarily conveyed by temporal modulations in the speech envelope in independent frequency bands, and any form of distortion that degrades those modulations will lead to a reduction in intelligibility. For example, both reverberation of a speech signal and background noise will tend to fill in the temporal dips of the directly

received speech envelope, degrading the speech modulation transfer. Similarly, if temporal dips in the speech envelope fall below the threshold of audibility in a frequency band, then intelligibility will be reduced.

A drawback of the AI, SII, and STI approaches is that any nonlinear aspects of processing by the human auditory system and degradations due to hearing loss can only be implemented in an ad hoc manner. Furthermore, it is now known that hearing impairment can occur without being reflected in the audiogram (i.e., the clinical measure of behavioral thresholds for hearing). Animal studies suggest that the loss of approximately 50-90% of auditory nerve (AN) fibers can occur without affecting the audiogram, and moderate sound exposures have the potential to cause such neural degeneration without producing any damage to the hair cells of the cochlea that determine behavioral thresholds (Liberman, 2016). However, because of the highly stochastic nature of AN responses at the single-fiber level, the loss of AN fibers is expected to degrade the representation of suprathreshold sounds such as speech (Lopez-Poveda and Barrios, 2013). These limitations in the acoustic-based metrics restrict their usefulness in understanding and quantifying how people perceive speech and, consequently, their utility for evaluating and improving devices such as hearing aids and cochlear implants. This has motivated the development of physiologically based intelligibility predictors by a number of different research groups.

Anatomy and Physiology of the Ear

Before jumping into the details of the physiological predictors, it is helpful to review the underlying auditory anatomy and physiology. **Figure 1** shows the anatomy of the human ear. The pinna and ear canal of the outer ear funnel sounds to the middle ear, which consists of the tympanic membrane (eardrum) and three tiny bones: the malleus, incus, and stapes. The footplate of the stapes lies atop the oval window, a membranous opening to the fluid-filled bony cochlea of the inner ear. The difference in size between the tympanic membrane and the oval window as well as the lever action of the middle ear bones produces a fairly good acoustic impedance match between the air in the ear canal and the fluids within the cochlea.

Figure 2 shows a cross section of the sensory structure in the cochlea, the organ of Corti, and **Figure 3** depicts how the acoustic vibrations conveyed to the oval window by the middle ear are transduced into neural "spikes" in the AN

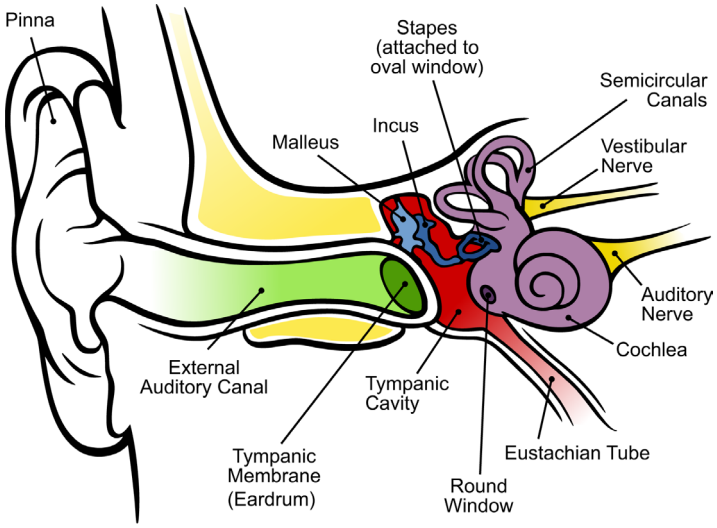

Figure 1. *Anatomy of the human ear. The pinna and external auditory canal form the outer ear. The middle ear consists of the tympanic membrane and the three tiny bones: the malleus, incus, and stapes. The cochlea is the sensory receptor organ of the inner ear, and the auditory nerve conveys the transduced acoustic information to the auditory pathways of the brain. Adapted from Chittka and Brockmann (2005) under the terms of the Creative Commons Attribution License © 2005 Chittka and Brockmann.*

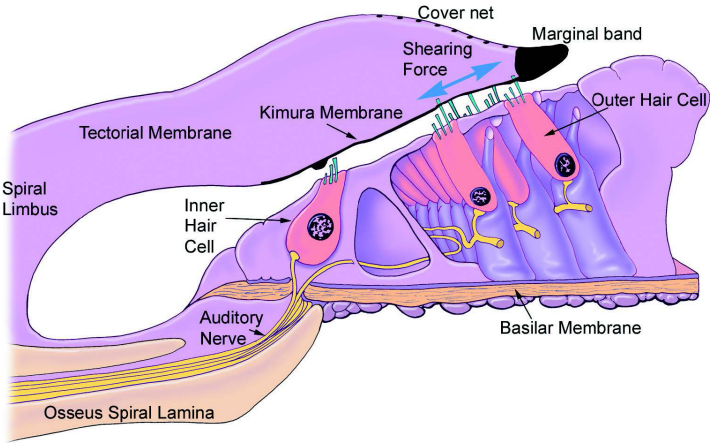

Figure 2. *Cross section of the organ of Corti in the cochlea. The basilar membrane (BM) spans the length of the fluid-filled cochlea, and its mechanical properties cause it to be tuned to different acoustic frequencies (see* ***Figure 3, bottom****). The organ of Corti, containing the sensory receptor cells known as inner hair cells (IHCs) and outer hair cells (OHCs), sits on top of the basilar membrane. Vertical vibration of the BM causes a shearing force to be applied to the hairlike cilia of the IHCs and OHCs, which, in turn, generates a transduction current and a subsequent change in the electrical potentials within the hair cells. The OHCs have an electromotile response (as described in the article by Brownell in this issue of* Acoustics Today*) that leads to a time-varying, nonlinear BM vibration pattern. The IHCs release a neurotransmitter that generates information-carrying "spikes" in the electrical potential of auditory nerve fibers. Reprinted with permission from Brownell (1997).*

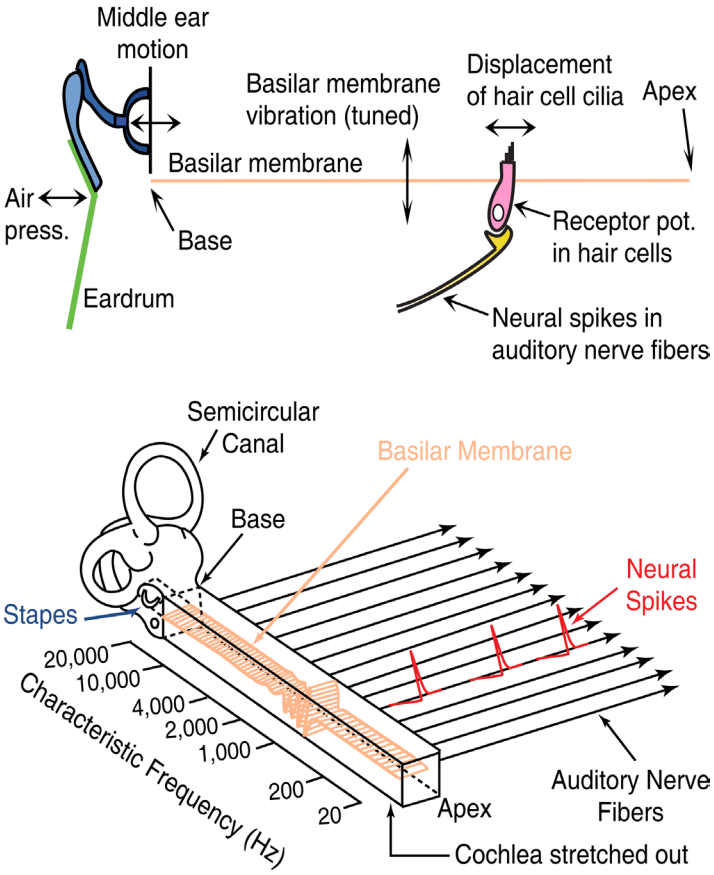

Figure 3. *Transduction of acoustic signals into neural activity by the ear.* ***Top:*** *illustration of how vibration of the eardrum (tympanic membrane) is transferred to the base of the cochlea by the bones of the middle ear. A pressure wave travels in the cochlear fluids from the base to the apex, with the mechanical tuning of the BM in the cochlea causing it to resonate to high acoustic frequencies at the base and low frequencies at the apex. Displacement of the IHC cilia by the shearing forces described in* ***Figure 2*** *leads to a change in the IHC potential, which triggers neurotransmitter release and subsequent generation of neural spikes in auditory nerve (AN) fibers.* ***Bottom:*** *further illustration of the frequency analysis performed by the cochlea. The resonant frequency at each point along the BM is referred to as the characteristic frequency (CF), and AN fibers inherit this frequency tuning because they connect to just one IHC. Thus, the AN conveys information about the acoustic stimulus by which fibers are responding as well as by the timing of the neural spikes. Adapted with permission from Sachs et al. (2002) © Biomedical Engineering Society.*

fibers by the organ of Corti. These spikes in the electrical potential of AN fibers are the basic information unit of the nervous system, and information about acoustic cues is encoded both by which AN fibers are spiking and by the timing of those spikes. The mechanical tuning of the basilar membrane within the cochlea leads to a "tonotopic map," where high frequencies generate responses in the base of the cochlea and lower frequencies generate responses further toward the apex. The electromotile action of the outer hair cells (OHCs) in the organ of Corti, as described in the article by Brownell in this issue of *Acoustics Today*, leads to the cochlea performing a time-varying, nonlinear time-frequency analysis of acoustic signals. It is this analysis that forms the basis of the physiologically based speech intelligibility predictors.

Framework for the Physiologically Based Predictions

Almost all neural speech intelligibility predictors are reference based, as illustrated in **Figure 4**. Such models create an "ideal" reference *(r)* response to a specific speech stimulus, that is, an unprocessed signal presented to a model of the normal auditory periphery at a conversational speech level in a quiet background. This forms a template of what the central auditory systems of the brain are expecting the AN activity to be for that particular stimulus. Such neural time-frequency representations are referred to a "neurograms." A comparison can then be made with the test case of a degraded *(d)* AN neurogram that differs from the ideal case because of modification of the acoustic stimulus and/or impairment of the auditory periphery.

One of the most widely used auditory-periphery models for speech intelligibility prediction is that of Zilany et al. (2014) and its predecessors. This model, illustrated in **Figure 5**, provides a high level of physiological detail of the transduction process in the ear, including the filtering of the middle ear, the nonlinear time-varying filtering of the cochlea in the inner ear (C1, C2, and control path filters), control of cochlear filtering by the OHCs, transduction of cochlear vibrations into electrical activity by the inner hair cells (IHCs), synaptic transmission, and generation of neural spiking activity in AN fibers. Alternative auditory-periphery models employ different degrees of physiological detail and accuracy to simplify the processing and increase computational efficiency (e.g., Elhilali et al., 2003; Jørgensen and Dau, 2011; Kates and Arehart, 2014), but models with greater amounts of physiological detail can generally provide more precise descriptions of neural speech coding and more detailed implementation of different forms of neural pathology, which is explored further in this article.

Although many intelligibility predictions are based directly on calculations comparing features of the reference and degraded AN neurograms, several metrics have taken inspiration from the STI (as well as other recent studies showing

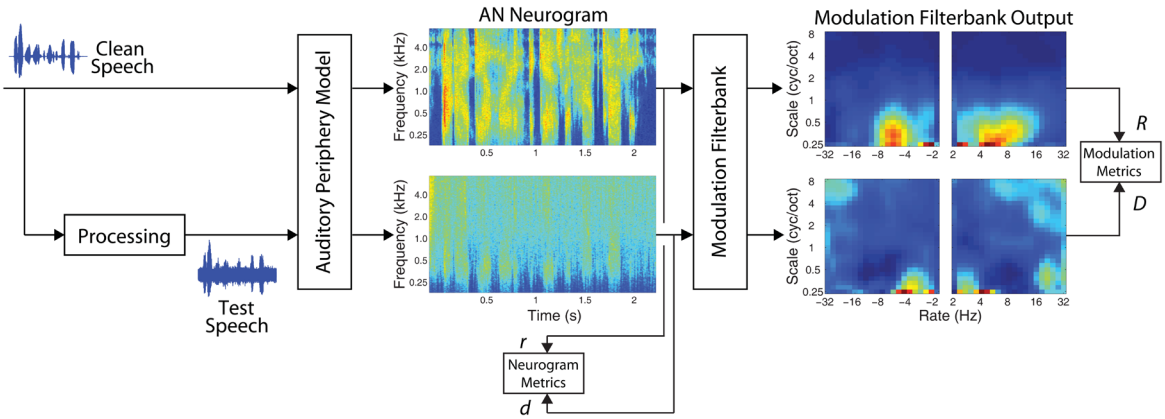

Figure 4. *Framework for reference-based prediction of speech intelligibility. A "reference" AN response "neurogram"(r;* ***middle top****) is obtained by presenting a clean speech signal to a normal-hearing AN model (see* ***Figure 5****) at a conversational speech level. A degraded AN neurogram (d;* ***middle bottom****) is generated for the same speech signal that may have undergone some form of processing to degrade the acoustic signal and/or for an AN model with some type of pathology. Neurogram-based metrics predict the intelligibility of the degraded speech by comparison of the r and d neurograms. Modulation-based metrics first pass the neurograms through a bank of modulation filters (see* ***Figure 6****) to generate corresponding R and D modulation representations for the reference and degraded speech, respectively, before computing an intelligibility prediction.*

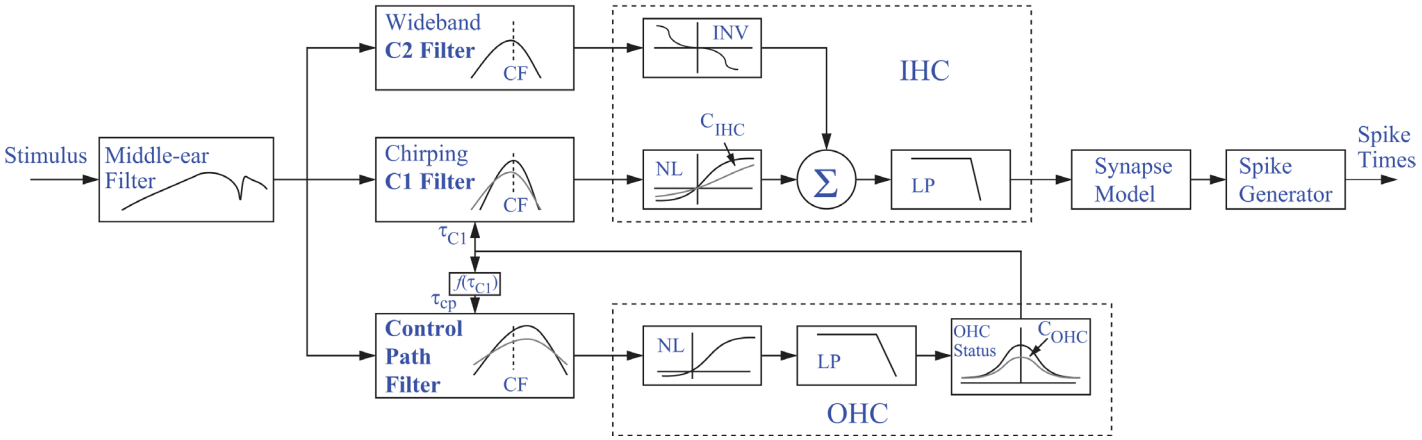

Figure 5. *Example of an AN fiber model used in physiologically based intelligibility predictors, providing a computational implementation of the transduction process described in* ***Figure 3****. The input to the model is an instantaneous pressure waveform of the acoustic stimulus impinging on the tympanic membrane, and the output is the set of spike times for a model AN fiber with a particular characteristic frequency (CF) in response to that input. The variables C_{OHC} and C_{IHC} shown respectively within the OHC and IHC blocks of the model, are scaling coefficients with values between 0 and 1 to indicate OHC and IHC health, respectively, at that CF in the cochlea. The variables τ_{C1} and τ_{cp} control the time-varying, nonlinear filtering in the signal and control paths, respectively. LP, low-pass filter; NL, static nonlinearity; INV, inverting nonlinearity; Σ, summation. Reprinted from Zilany and Bruce (2006) with permission from the Acoustical Society of America © 2006.*

the importance of speech modulations for intelligibility) and incorporate modulation filter banks (Elhilali et al., 2003; Zilany and Bruce, 2007; Jørgensen and Dau, 2011). As seen in **Figure 4,** ***right***, the reference (*r*) and degraded (*d*) AN neurograms can be passed through such a modulation filter bank to produce corresponding reference (*R*) and degraded (*D*) modulation spectrum representations. **Figure 6** illustrates the spectrotemporal receptive fields (STRFs) of a widely used filter bank that considers joint time-frequency modulations (Elhilali et al., 2003; Zilany and Bruce, 2007). Other predictors consider one or more banks of filters that analyze temporal modulations within each frequency band (e.g., Jørgensen and Dau, 2011).

Note that in contrast to the widely used reference-based approaches, it is also possible to obtain "reference-free" predictions of speech based on how the output of an auditory model responding to a test speech signal varies in certain statistics from the average neural response to speech in general rather than a reference version of the specific speech signal (e.g., Hossain et al., 2016).

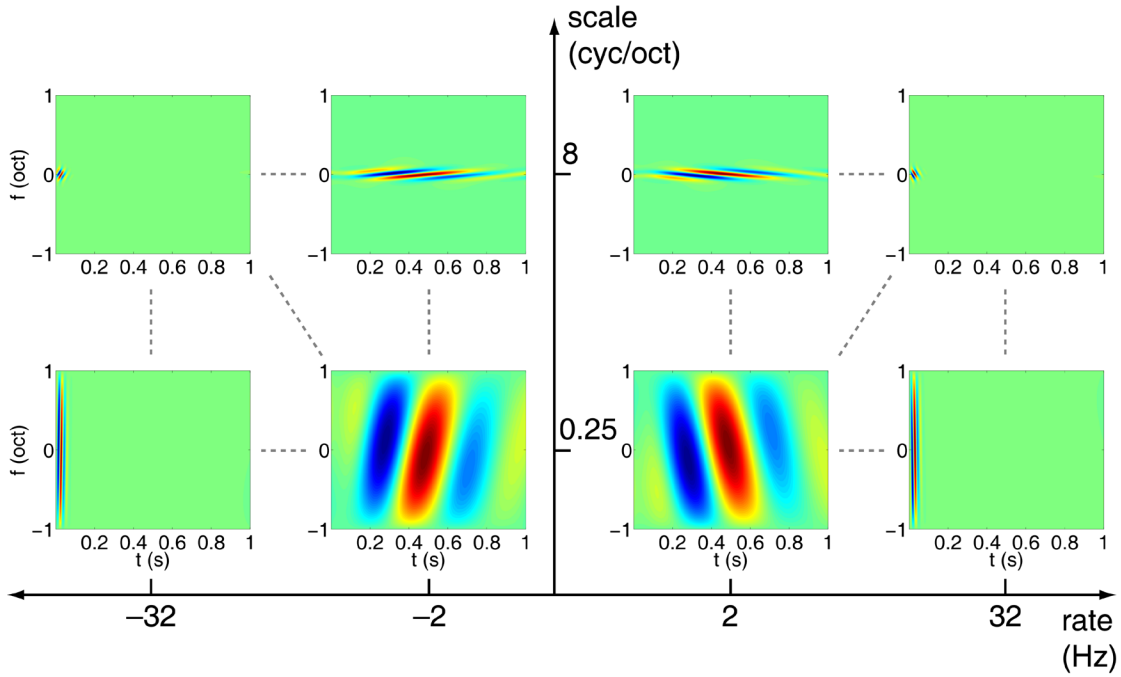

Figure 6. *Example of a modulation filter bank. Each plot shows a spectrotemporal receptive field (STRF) for a different modulation filter, showing time-frequency regions of excitation (**red**) and regions of inhibition (**blue**). Filters with low temporal modulation rates (close to the y-axis) prefer slower temporal fluctuations in the AN neurogram and filters with higher modulation rates prefer faster temporal oscillations. Filters with low spectral modulation scales (close to the x-axis) prefer broad spectral peaks and valleys in the AN neurogram and filters with high spectral modulation scales prefer more closely spaced spectral variations. Positive and negative temporal modulation rates correspond to downward and upward frequency sweeps, respectively, in the STRF. cyc/oct, Cycles/octave. Adapted from Bruce and Zilany (2007).*

Examples of Predictor Computation Approaches

An example of how intelligibility can be predicted directly from AN neurograms is the Neurogram SIMilarity (NSIM) metric developed by Hines and Harte (2012). A 3×3 time-frequency kernel is swept over the neurograms, and for each of these kernels, an NSIM value is calculated according to

$$\mathrm{NSIM}(\mathrm{r},\mathrm{d}) = \left(\frac{2\mu_\mathrm{r}\mu_\mathrm{r} + C_1}{\mu_\mathrm{r}^2 + \mu_\mathrm{d}^2 + C_1}\right)^{\alpha} \cdot \left(\frac{2\sigma_\mathrm{r}\sigma_\mathrm{d} + C_2}{\sigma_\mathrm{d}^2 + \sigma_\mathrm{d}^2 + C_2}\right)^{\beta} \cdot \left(\frac{\sigma_\mathrm{rd} + C_3}{\sigma_\mathrm{r}\sigma_\mathrm{d} + C_3}\right)^{\gamma} \qquad (1)$$

where the first term compares the "luminance" of the neurogram images (i.e., the average intensity [μ_x] of each kernel), the second term compares the visual "contrast" of the neurograms (i.e., the standard deviation [σ_x] in each kernel), and the third term assesses the "structural" relationship between the two neurograms (conveyed as the Pearson product-moment correlation coefficient [σ_{xy}] in each kernel). The coefficients α, β, and γ can take values between 0 and 1; for the speech material tested thus far in the literature, the optimal values for predicting speech intelligibility are $\alpha = \gamma \approx 1$ and $\beta \approx 0$ (Hines and Harte, 2012; Bruce et al., 2013). The constants C_1, C_2, and C_3 regularize the calculation in cases where the denominator may approach a value of 0. The overall NSIM value is obtained by averaging the values computed for each 3×3 kernel, with a theoretical range of 0 (totally unintelligibility) to 1 (completely intelligible).

An important feature of the NSIM is that the time resolution of the AN neurogram can be adjusted such that it only takes into account the mean-rate (MR) information in the AN response (**Figure 7, *middle***) or that it also includes fine-timing (FT) information about neural spike times (**Figure 7, *bottom***).

Alternative neurogram-based metrics include neural correlation (Bondy et al., 2004; Christiansen et al., 2010), shuffled correlograms (Swaminathan and Heinz, 2012), envelope correlation coupled with audio coherence (Kates and Arehart, 2014), an orthogonal polynomial measure (Mamun et al., 2015), and bispectrum analysis (Hossain et al., 2016).

An example of a modulation-based predictor is the spectrotemporal modulation index (STMI) developed by Elhilali et al. (2013), in which a normalized distance between the reference and degraded spectrotemporal modulation spectra is computed according to

$$\mathrm{STMI} = 1 - \frac{\|R - D\|^2}{\|R\|^2} \qquad (2)$$

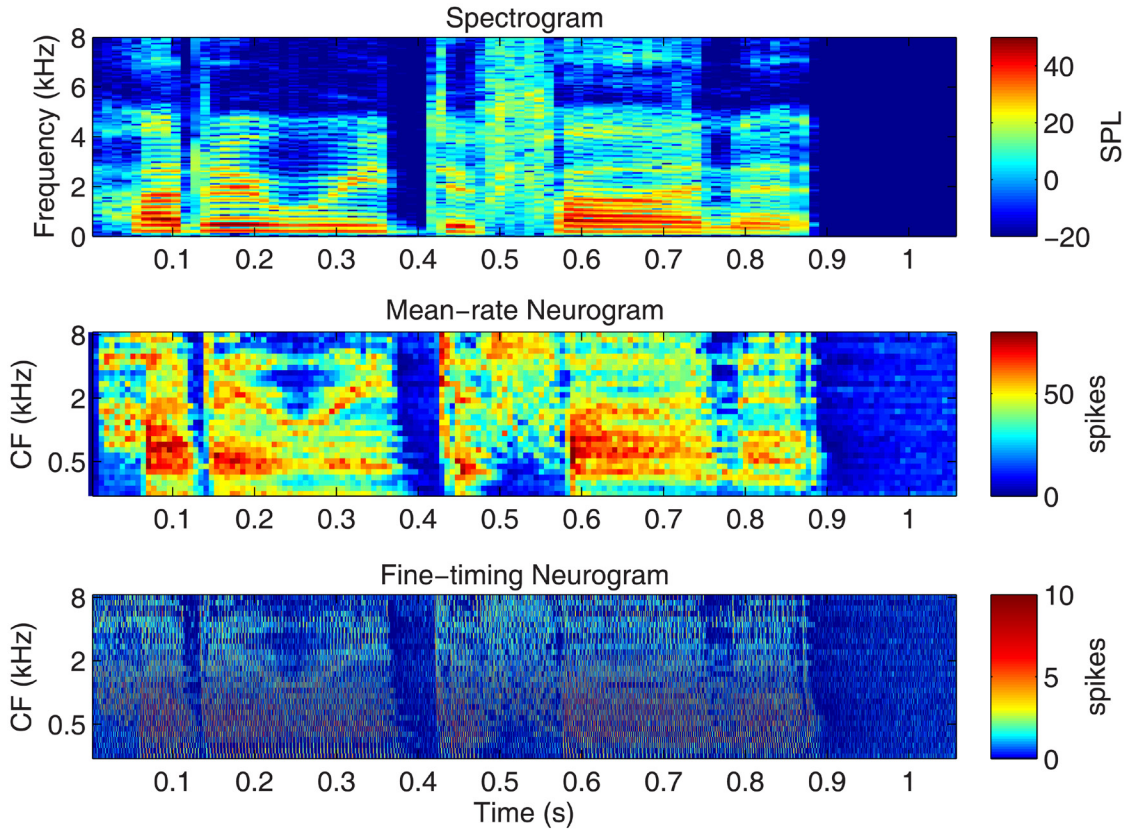

Figure 7. *Neural representation of a reference speech signal, the sentence "How do we define it?", presented to a normal-hearing AN model at 65 dB sound pressure level (SPL) in a quiet background.* ***Top:*** *acoustic spectrogram for this sentence.* ***Middle:*** *mean-rate (MR) neurogram for a population of AN fibers with CFs between 0.25 and 8 kHz and a time resolution of 6.4 ms.* ***Bottom:*** *fine-timing (FT) neurogram for the same set of AN fibers but with a time resolution of 0.16 ms.*

Similar to the NSIM, the STMI theoretically approaches a value of 1 for perfect intelligibility. In situations where the degradations in the test case reduce the neural modulations, the STMI will approach a value of 0. However, if the modulations in the test case are greater than those of the reference case, it is possible for the STMI to go negative. To avoid this, the difference ($R - D$) for each time, the characteristic frequency (CF), the temporal modulation rate, and the spectral modulation scale combination can be set to 0 for negative values, as done by Zilany and Bruce (2007). Some alternative modulation-based metrics instead compute an SNR in the modulation domain to mitigate this problem (Jørgensen and Dau, 2011; Rallapalli and Heinz, 2016).

Example Predictions

To illustrate how background noise can degrade the AN neurogram, the simulations in **Figure 7** were repeated with added background white Gaussian noise at an SNR of 0 dB. **Figure 8** shows that such a high level of noise can fill in the quiet gaps of the speech signal, but the higher energy, low-frequency acoustic features are relatively well preserved in the MR response (***middle***), and neural synchrony to these features can also be observed in the FT neurogram (***bottom***).

As mentioned in the **Introduction**, AN fiber degeneration would also be expected to degrade the overall neural representation, an effect that cannot be captured by the acoustic-based predictors. This is illustrated in **Figure 9**, where the speech stimulus is presented in quiet (as in **Figure 7**) but with only 30% survival of AN fibers. Here the patterns of MR and spike-timing information in the MR and FT neurograms are more similar to those in **Figure 7**, but the overall response in **Figure 9** is more subject to neural noise because of the reduced number of AN fibers.

Figure 10 shows how the NSIM predictions are affected by different SNRs (***left***) and different amounts of neural survival (***right***). In both cases, a monotonic relationship is observed, and the empirical maximal values are less than the theoretical value of 1 because of the stochastic AN fiber activity. The maximal values for the FT NSIM are lower than those of the MR NSIM because the larger time bins used for the latter are more effective at averaging out the neural noise. However, at low SNRs, the FT NSIM continues to show reductions, with a decreasing SNR when the MR NSIM reached its lower asymptote, indicating that some spike-timing information about the sentence may be preserved at very low SNRs, even

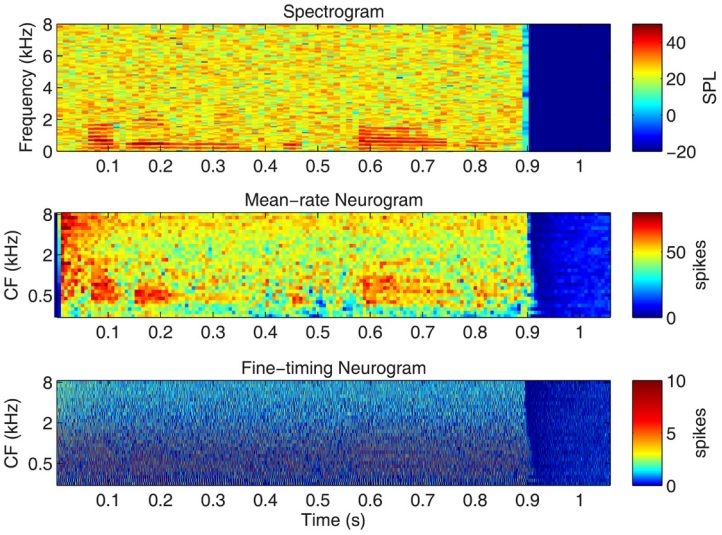

Figure 8. *Spectrogram (**top**) and neurograms (**middle** and **bottom**) for the sentence "How do we define it?" presented to a normal-hearing AN model at 65 dB SPL in the presence of background white Gaussian noise, also at 65 dB SPL. Plotting conventions are the same as in **Figure 4**.*

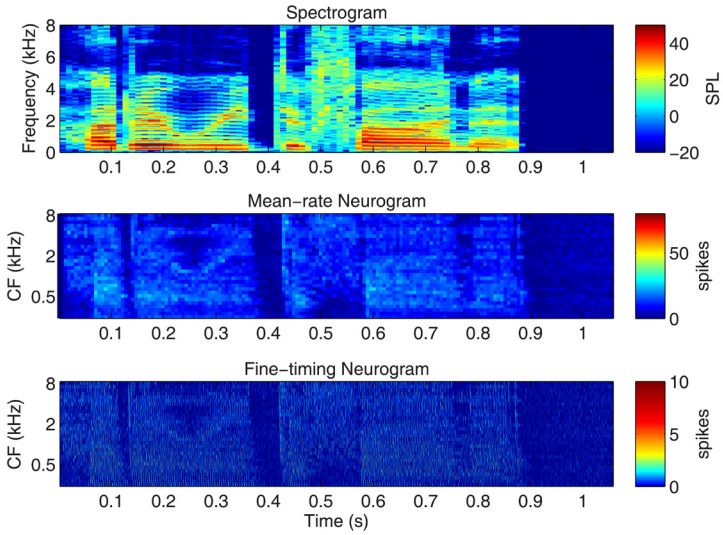

Figure 9. *Spectrogram (**top**) and neurograms (**middle** and **bottom**) for the sentence "How do we define it?" at 65 dB SPL presented to an AN model with only 30% neural survival. Plotting conventions are the same as in **Figure 4**.*

when the MR cues have been totally lost. Similarly, the FT NSIM does not drop as rapidly with decreasing neural survival as does the MR NSIM, suggesting that the spike-timing representation may be more robust to loss of AN fibers.

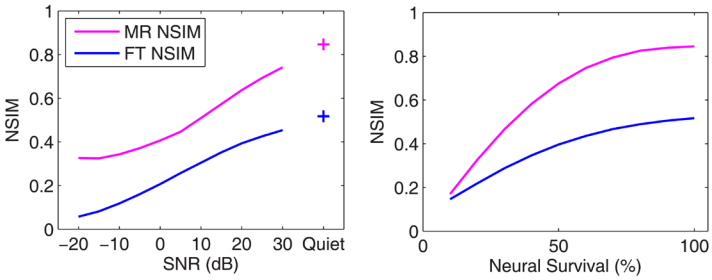

Figure 10. *MR and FT Neurogram SIMilarity (NSIM) metric predictions for the sentence "How do we define it?" as a function of the signal-to-noise ratio (SNR) for background white Gaussian noise (**left**) and as a function of AN fiber survival (**right**).*

Lessons Learned (and To Be Learned) from Physiologically Based Predictors

In general, physiologically based metrics have performed as well as or better than the traditional acoustic-based metrics in quantitative predictions of perceptual data (e.g., Bondy et al., 2004; Jørgensen and Dau, 2011; Christiansen et al., 2010; Bruce et al., 2013). This indicates that adding physiological detail is not problematic for cases where the acoustic-based metrics perform reasonably well and that the neural-based metrics can indeed overcome some of the shortcomings of the acoustic predictors. However, there is a range of different metrics that have been developed, as reviewed in this article, but only limited head-to-head comparisons have been performed between the physiological predictors (e.g., Bruce et al., 2013; Chabot-Leclerc et al., 2014). An important future area of research is to conduct more rigorous comparisons of the different predictor for multiple sets of speech intelligibility data to determine which approach is best in general.

In addition, studies using these predictors have given insight into the neural coding of speech features. In general, the MR representation of speech envelope cues (i.e., slower temporal modulations) appears to be the dominant neural representation in most situations, but spike-timing cues may also contribute additional information in adverse conditions such as low SNRs (e.g., Swaminathan and Heinz, 2012; Bruce et al., 2013). These conclusions have important consequences for how hearing aids and cochlear implants encode the acoustic features of speech, and physiologically based intelligibility predictors should be important tools in the improvement of such devices for the hearing impaired (Sachs et al., 2002).

Biosketch

Ian C. Bruce completed his bachelor's degree in electrical and electronic engineering at the University of Melbourne, Melbourne, VIC, Australia, and his PhD at the university's Bionic Ear Institute. In between, he was a research and teaching assistant at the University of Technology in Vienna, Austria. He did a postdoctoral fellowship in the Department of Biomedical Engineering at the Johns Hopkins University in Baltimore, MD, before joining the faculty at McMaster University in Hamilton, ON, Canada, in 2002. He is currently an associate professor in electrical and computer engineering and is engaged in interdisciplinary activities in neuroscience, biomedical engineering, psychology, and music cognition.

References

Allen, J. B. (1996). Harvey Fletcher's role in the creation of communication acoustics. *The Journal of the Acoustical Society of America* 99, 1825-1839.

American National Standards Institute (ANSI). (1997). ANSI S3.5-1997. *Methods for Calculation of the Speech Intelligibility Index*. American National Standards Institute, New York.

Bondy, J., Bruce, I. C., Becker, S., and Haykin, S. (2004). Predicting speech intelligibility from a population of neurons. In Thrun, S., Saul, L., and Schölkopf, B. (Eds.), *Advances in Neural Information Processing Systems* 16. MIT Press, Cambridge, MA, pp. 1409-1416.

Brownell, W. E. (1997). How the ear works - Nature's solutions for listening. *Volta Review* 99, 9-28.

Bruce, I. C., and Zilany, M. S. A. (2007). Modelling the effects of cochlear impairment on the neural representation of speech in the auditory nerve and primary auditory cortex. In Dau, T., Buchholz, J., Harte, J. M., and Christiansen, T. U. (Eds.), *Auditory Signal Processing in Hearing-Impaired Listeners, International Symposium on Audiological and Auditory Research (ISAAR)*, Danavox Jubilee Foundation, Nyborg, Denmark, August 23-25, 2007, pp. 1-10.

Bruce, I. C., Léger, A. C., Moore, B. C. J., and Lorenzi, C. (2013). Physiological prediction of masking release for normal-hearing and hearing-impaired listeners. *Proceedings of Meetings on Acoustics: ICA 2013 Montreal*, Montreal, QC, Canada, June 2-7, 2013, vol. 19, 050178. doi:10.1121/1.4799733.

Chabot-Leclerc, A., Jørgensen, S., and Dau, T. (2014). The role of auditory spectro-temporal modulation filtering and the decision metric for speech intelligibility prediction. *The Journal of the Acoustical Society of America* 135, 3502-3512. doi:10.1121/1.4873517.

Chittka, L., and Brockmann, A. (2005). Perception space—The final frontier. *PLoS Biology* 3, e137. doi:10.1371/journal.pbio.0030137.

Christiansen, C., Pedersen, M. S., and Dau, T. (2010). Prediction of speech intelligibility based on an auditory preprocessing model. *Speech Communication* 52, 678-692. doi:10.1016/j.specom.2010.03.004.

Elhilali, M., Chi, T., and Shamma, S. A. (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication* 41, 331-348. doi:10.1016/S0167-6393(02)00134-6.

Fletcher, H., and Galt, R. (1950). Perception of speech and its relation to telephony. *The Journal of the Acoustical Society of America* 22, 89-151.

French, N. R., and Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America* 19, 90-119.

Hines, A., and Harte, N. (2012). Speech intelligibility prediction using a Neurogram Similarity Index Measure. *Speech Communication* 54, 306-320. doi:10.1016/j.specom.2011.09.004.

Hornsby, B. W. Y. (2004). The Speech Intelligibility Index: What is it and what's it good for? *The Hearing Journal* 57, 10-17.

Hossain, M. E., Jassim, W. A., and Zilany, M. S. A. (2016). Reference-free assessment of speech intelligibility using bispectrum of an auditory neurogram. *PLoS ONE* 11, e0150415. doi:10.1371/journal.pone.0150415.

Jørgensen, S., and Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America* 130, 1475-1487. doi:10.1121/1.3621502.

Kates, J. M., and Arehart, K. H. (2014). The Hearing-Aid Speech Perception Index (HASPI). *Speech Communication* 65, 75-93. doi:10.1016/j.specom.2014.06.002.

Liberman, M. C. (2016). Noise-induced hearing loss: Permanent versus temporary threshold shifts and the effects of hair cell versus neuronal degeneration. In Popper, A. N., and Hawkins, A. (Eds.), *The Effects of Noise on Aquatic Life II*, Springer-Verlag, New York, pp. 1-7. doi:10.1007/978-1-4939-2981-8_1.

Lopez-Poveda, E. A., and Barrios, P. (2013). Perception of stochastically undersampled sound waveforms: A model of auditory deafferentation. *Frontiers in Neuroscience* 7, 1-13. doi:10.3389/fnins.2013.00124.

Mamun, N., Jassim, W. A., and Zilany, M. S. A. (2015). Prediction of speech intelligibility using a neurogram orthogonal polynomial measure (NOPM). *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23, 760-773. doi:10.1109/TASLP.2015.2401513.

Rallapalli, V. H., and Heinz, M. G. (2016). Neural spike-train analyses of the speech-based envelope power spectrum model: Application to predicting individual differences with sensorineural hearing loss. *Trends in Hearing* 20, 1-14. doi:10.1177/2331216516667319.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2006). Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. *The Journal of the Acoustical Society of America* 120, 3988-3997. doi:10.1121/1.2358008.

Sachs, M. B., Bruce, I. C., Miller, R. L., and Young, E. D. (2002). Biological basis of hearing aid design. *Annals of Biomedical Engineering* 30, 157-168. doi:10.1114/1.1458592.

Steeneken, H. J. M., and Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America* 67, 318-326.

Swaminathan, J., and Heinz, M. G. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *Journal of Neuroscience* 32, 1747-1756. doi:10.1523/JNEUROSCI.4493-11.2012.

Zilany, M. S. A., and Bruce, I. C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *The Journal of the Acoustical Society of America* 120, 1446-1466. doi:10.1121/1.2225512.

Zilany, M. S. A., and Bruce, I. C. (2007). Predictions of speech intelligibility with a model of the normal and impaired auditory-periphery. *Proceedings of 3rd International IEEE/EMBS Conference on Neural Engineering*, Kohala Coast, HI, May 2-5, 2007, pp. 481-485. doi:10.1109/cne.2007.369714.

Zilany, M. S. A., Bruce, I. C., and Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America* 135, 283-286. doi:10.1121/1.4837815.