

## Supplemental Material S1.

### *Individualized Amplification of Speech*

For the OHI listeners, source speech signals were amplified using an individualized gain profile such that third-octave speech levels would be at least 10 dB above the listener’s FRESH-noise audiometric thresholds at frequencies up to 4 kHz (Humes & Coughlin, 2009). This was achieved by first measuring the third-octave frequency response of EARTONE 3A insert earphones (to be used for stimulus presentation) in a 2-cc coupler (Larson Davis model AEC203). A 30 s Gaussian noise calibration signal was filtered into third-octave bands using a sixth-order, class 1 digital filter bank with center frequencies from 125–8000 Hz and magnitude attenuation limits as specified in ANSI S1.11-2004 (Acoustical Society of America, 2004). Each time-domain, band-limited signal was then normalized to a common root-mean-square (RMS) level that would produce a 2-cc coupler level of ~75 dB SPL for the broadband signal. The actual 2-cc coupler level (dB SPL) produced by each band-limited signal was recorded on a Larson Davis System 824 sound level meter. These values were converted to dB HL by subtracting the reference equivalent threshold SPLs of the EARTONE 3A headset (Acoustical Society of America, 2004).

Speech signals were presented diotically to ONH listeners at an overall level of 75 dB SPL. For OHI listeners, speech signals were first set to an overall level of 75 dB SPL and then separated into third-octave bands using the same filter bank as used for the calibration noise. The third-octave speech levels were converted to dB HL based on the levels produced by the calibration noise in dB HL. The listener’s audiometric thresholds were interpolated over frequency (linear interpolation on the log scale) to match the center frequencies of the third-octave filter bank, and the RMS level in any band for which the speech level did not exceed the listener’s audiometric threshold by at least 10 dB HL was scaled to produce a third-octave level exactly 10 dB above threshold. The individual bands were then summed to produce an amplified broadband signal. This process was performed separately for each stimulus presentation and individualized to the audiogram of each ear.

### *Bubbles Filtering*

The bubbles filtering procedure using phase gradient heap integration (PGHI) was implemented in MATLAB R2015b (The MathWorks, Inc.) using the Phase Retrieval (“phaseret”) toolbox v0.2.1 and the Large Time/Frequency Analysis (“lftfat”) toolbox v2.2.1 (Průša *et al.*, 2014). For each two-talker mixture, a spectrogram was obtained via the discrete gabor transform using a Gaussian window with a frequency resolution of 16.3 Hz and a hop time of 1 ms. The log-magnitude spectrogram was transformed into the modulation power spectrum (MPS) domain via two-dimensional fast Fourier transform (2D-FFT) where the resulting representation had a pixel-by-pixel resolution of 0.5 Hz in the temporal modulation dimension and 0.1 cyc/kHz in the spectral modulation dimension. A binary bubbles filter was applied to MPS pixels encoding spectral modulations  $\leq 15$  Hz and temporal modulations  $< 50$  Hz ( $151 \times 99$  pixels). The bubbles filter was applied mirror-symmetrically to regions of the MPS encoding upward- and downward-sweeping spectrotemporal modulations (Venezia *et al.*, 2016, 2019). The bubbles filter was created by generating an all-zero image of size  $151 \times 99$  pixels, setting a randomly-selected subset of pixels (equal to the number of bubbles, chosen adaptively for each

listener) to a value of one, smoothing with a Gaussian filter with  $\sigma = 5$  pixels (0.5 cyc/kHz) in the spectral modulation dimension and  $\sigma = 4$  pixels (2 Hz) in the temporal modulation dimension, and binarizing the resulting image using a cutoff value of 0.1. Finally, the binary bubbles filter was multiplied with the MPS, a filtered magnitude spectrogram was obtained by inverse 2D-FFT, and a filtered waveform was obtained by PGHI.

### **Multinomial Logistic Regression Procedure**

Model fitting was carried out using the *glmnet* package v2.0-16 in R v3.4.4. For a response variable with  $K$  categories, the multinomial logistic regression problem is characterized by *glmnet* as follows (Hastie & Qian, 2014):

$$\Pr(G = k | X = x) = \frac{e^{\beta_{0k} + \beta_k^T x}}{\sum_{l=1}^K e^{\beta_{0l} + \beta_l^T x}}$$

where  $k$  is a given response category,  $x$  is the predictor matrix,  $\beta$  is a  $p \times k$  matrix of regression coefficients (weights),  $p$  is the number of predictors in  $x$ ,  $\beta_0$  is a length- $p$  vector of regression coefficients belonging to a reference category, and  $l$  is an indicator variable summing over the  $k$  response categories. The model estimates  $\Pr(G = k | X = x)$  as an  $n \times K$  matrix of probabilities, where  $n$  is the number of observations and each row (observation) sums to one. The value in each column indicates the probability that an observation belongs to the  $k$ th response category given the set of predictor values,  $x$ . The best-fitting set of weights,  $\beta$ , is determined by minimizing the model deviance:

$$-2 \sum_{i=1}^n \sum_{j=1}^K y_{ij} \log \frac{\Pr(G = j | X = x_i)}{y_{ij}}$$

where  $i$  indexes the observation number,  $j$  indexes the response category,  $y_{ij}$  is the  $j$ th value of a length- $K$  binary vector indicating which response category observation  $i$  belongs to, and  $x_i$  is the length- $p$  vector of predictor values for observation  $i$ . This reduces to twice the sum over observations of the multi-log loss, which is the minus log-probability assigned by the model to the actual response category. An L2 (ridge) penalty term approximately equal to a constant,  $\lambda$ , times the squared magnitude of the regression coefficients,  $\beta$ , was included in the likelihood function of the regression model. The constant  $\lambda$  is thus a hyper-parameter of the model and must be tuned by cross-validation. The reader is referred to Hastie and Qian (2014) for a complete description of the multinomial likelihood function in *glmnet*.

In the context of the current experiment, the predictors,  $x$ , are the trial-by-trial, bubbles-filter feature vectors, and each column of the matrix of regression coefficients,  $\beta$ , is a classification image (CIImg) once reshaped to the original dimensions of  $151 \times 40$  pixels. In fact, multinomial CIImgs must be interpreted relative to a reference category ( $\beta_0$ ), which is the ‘incorrect’ category. Therefore, we obtain three CIImgs from the multinomial regression model: a color+number CIImg, a color-only CIImg, and number-only CIImg. These CIImgs are obtained by taking the difference between  $\beta_k$  for a given response category (color+number, color only, number only) and  $\beta_0$  for the reference category (incorrect). We also obtain predictions from the fitted model indicating the probability that a given observation belongs to each of the possible response categories: color+number, color only, number only, and incorrect. The predictive

performance of the model is characterized by the mean multi-log loss across the set of observations for which predictions were generated.

For model fitting, the data from a given participant were split into pseudo-randomly selected training (75% of trials) and test (25% of trials) sets balanced for the distribution of response types. The value of hyper-parameter  $\lambda$  was then optimized on the training set using 10-fold cross-validation. The search space for  $\lambda$  included powers of two from 0-9 in steps of 0.1. The optimal value of  $\lambda$  was taken to be the largest such value whose (across-fold) average deviance was within one standard error of the minimum average deviance over the entire search space (Hastie et al., 2009). The predictor variables were centered and scaled separately for each cross-validation fold to avoid information leak. Once an optimal  $\lambda$  value was obtained for each participant, the sample-wide optimal lambda value was obtained as the average of the individual-participant optimal lambda values. A sample-wide optimal  $\lambda$  was obtained so that the regression coefficients (CImgs) would be equally smooth (larger  $\lambda$  = greater smoothing) across participants. The optimal values of lambda were not significantly different between ONH and OHI listener groups ( $t_{18} = 1.49$ ,  $p = 0.15$ ). Using the sample-wide optimal lambda value, the multinomial regression model was re-fitted to the entire training data set of each participant. Predictor variables were again centered and scaled. From these fully trained participant-level regression models, CImgs were obtained for each participant and the model’s predictive performance was calculated as the multi-log loss of the model predictions on each participant’s test data set. When obtaining predictive performance, the predictor variables in the test set were centered and scaled using the mean and standard deviation obtained from training set to avoid information leak.

## References

- Acoustical Society of America. (2004). *American National Standard Specification for Octave-band and Fractional-octave-band Analog and Digital Filters: ANSI S1.11-2004*.
- Hastie, T., & Qian, J. (2014). *Glmnet vignette*. [http://www.web.stanford.edu/~hastie/Papers/Glmnet\\_Vignette.pdf](http://www.web.stanford.edu/~hastie/Papers/Glmnet_Vignette.pdf)
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- Humes, L. E., & Coughlin, M. (2009). Aided speech-identification performance in single-talker competition by older adults with impaired hearing. *Scandinavian Journal of Psychology* 50(5), 485–494. <https://doi.org/10.1111/j.1467-9450.2009.00740.x>
- Průša, Z., Søndergaard, P. L., Holighaus, N., Wiesmeyr, C., & Balazs, P. (2014). The Large Time-Frequency Analysis Toolbox 2.0. In M. Aramaki, O. Derrien, R. Kronland-Martinet, & S. Ystad (Eds.), *Sound, music, and motion, lecture notes in computer science* (pp. 419–442). Springer.
- Venezia, J. H., Hickok, G., & Richards, V. M. (2016). Auditory “bubbles”: Efficient classification of the spectrotemporal modulations essential for speech intelligibility. *The Journal of the Acoustical Society of America*, 140(2), 1072–1088. <https://doi.org/10.1121/1.4960544>
- Venezia, J. H., Martin, A.-G., Hickok, G., & Richards, V. M. (2019). Identification of the spectrotemporal modulations that support speech intelligibility in hearing-impaired and normal-hearing listeners. *Journal of Speech, Language, and Hearing Research*, 62(4), 1051–1067. [https://doi.org/10.1044/2018\\_JSLHR-H-18-0045](https://doi.org/10.1044/2018_JSLHR-H-18-0045)