The Noisy Encoding of Disparity Model of the McGurk Effect

John F. Magnotti[1] and **Michael S. Beauchamp**[1]

[1]Department of Neurobiology & Anatomy, University of Texas Medical School at Houston

Abstract

In the McGurk effect, incongruent auditory and visual syllables are perceived as a third, completely different syllable. This striking illusion has become a popular assay of multisensory integration for individuals and clinical populations. However, there is enormous variability in how often the illusion is evoked by different stimuli and how often the illusion is perceived by different individuals. Most studies of the McGurk effect have used only one stimulus, making it impossible to separate stimulus and individual differences. We created a probabilistic model to separately estimate stimulus and individual differences in behavioral data from 165 individuals viewing up to 14 different McGurk stimuli. The noisy encoding of disparity (NED) model characterizes stimuli by their audiovisual disparity and characterizes individuals by how noisily they encode the stimulus disparity and by their disparity threshold for perceiving the illusion. The model accurately described perception of the McGurk effect in our sample, suggesting that differences between individuals are stable across stimulus differences. The most important benefit of the NED model is that it provides a method to compare multisensory integration across individuals and groups without the confound of stimulus differences. An added benefit is the ability to predict frequency of the McGurk effect for stimuli never before seen by an individual.

Keywords

McGurk effect; individual differences; speech perception; multisensory integration

When we watch someone speaking, we combine visual speech information from the talker's mouth with the auditory speech information from the talker's voice to increase recognition speed and accuracy. When the auditory and visual speech are incongruent, combining the two information sources can lead to a fused percept distinct from both the auditory and visual speech. This illusion, termed the McGurk effect, has been widely used to understand temporal constraints on audiovisual binding (Munhall, Gribble, Sacco, & Ward, 1996; Stevenson, Zemtsov, & Wallace, 2012) and to characterize differences in multisensory integration across different individuals (Green, Kuhl, Meltzoff, & Stevens, 1991; MacDonald & McGurk, 1978; Nath & Beauchamp, 2012; Sekiyama, 1997), different ages in the lifespan (Tremblay et al., 2007), and different clinical populations (Irwin, Tornatore, Brancazio, & Whalen, 2011; Mongillo et al., 2008; Woynaroski et al., 2013).

Correspondence should be sent to: John F. Magnotti, Ph.D. or Michael S. Beauchamp, Ph.D., 6431 Fannin St. Suite G.550, Houston, TX 77030, john.magnotti@gmail.com; michael.s.beauchamp@uth.tmc.edu.

Using the McGurk effect to measure multisensory integration is advantageous because it can be measured simply: a few repetitions of a simple language stimulus. However, the McGurk effect is not homogenous, and shows large differences across both different stimuli and different individuals. In the original report, McGurk and MacDonald (1976) found that 98% of adults perceived the illusion for one stimulus (reporting the fused "da" percept when auditory speech "ba" was dubbed onto visual speech "ga") while only 81% perceived the illusion with a different stimulus (reporting the fused "ta" percept when an auditory "pa" was dubbed onto a visual "ka"). Across individuals, some participants almost always perceive the McGurk effect while others rarely do (Nath & Beauchamp, 2012; Gentiulcci & Cattaneo, 2005; Stevenson, Zemtsov, & Wallace, 2012).

We created a model of the McGurk effect to provide a rational description of these inter-stimulus and inter-individual differences. If the same individual perceives different stimuli differently, we assume that this reflects stimulus differences; if different individuals perceive the same stimulus differently, we assume that this reflect individual differences.

Recent evidence suggests that individual differences in McGurk perception are linked to differences in brain activity in the superior temporal sulcus (STS). Nath and Beauchamp (2012) used blood-oxygen level dependent functional magnetic resonance imaging (BOLD fMRI) to show that the response amplitude in the left superior temporal sulcus (STS) was 50% higher for individuals that frequently perceived the McGurk effect, compared to individuals that only infrequently perceived the illusion. Beauchamp, Nath, and Pasalar (2010) showed that interrupting activity in the left STS using transcranial magnetic stimulation reduced illusion perception by as much as 50%, while Keil, Muller, Ihssen, and Weisz (2012) showed that the ongoing activity in STS regions just prior to presentation of an incongruent stimulus predicted perception of the McGurk effect. Therefore, another goal of the model is to use subject-specific parameters that reflect our current understanding of the neural basis of multisensory integration and that have the potential to be relatable to neuroimaging measures of human brain function.

Studies of neural processing estimate a measure of neural activity that contains a single parameter, such as the neuronal firing rate in single neuron studies or the amplitude of activity in BOLD fMRI studies. To allow for comparison with brain activity measures, our model contains a single parameter describing each individual's propensity to integrate auditory and visual information.

Studies of neural processing may also estimate a measure of neural variability. Even when the same stimulus is presented repeatedly, the neurons encoding it show variability in their firing patterns (Mainen & Sejnowski, 1995). This variability represents a fundamental limitation on the representational precision of any stimulus. To account for this noise, the brain is thought to use Bayesian inference: given a noisy sensory observation, the brain infers the most likely state of the world (Angelaki, Gu, & DeAngelis, 2009; Knill & Pouget, 2004; Ma, Zhou, Ross, Foxe, & Parra, 2009). To aid in comparison with studies of brain activity (which typically report a single measure of neural variability) our model estimates a single sensory noise parameter for each participant.

Different stimuli are differently able to elicit the McGurk effect (Jiang & Bernstein, 2011; McGurk & MacDonald, 1976). Because the stimuli used in the original McGurk and MacDonald study are not available, different studies commonly use different stimuli. Comparing the raw proportion of McGurk perception between individuals tested with different stimuli that likely differ in their efficacy is problematic. The NED model characterizes each stimulus with a single parameter termed audiovisual disparity, which describes the difference between the auditory and visual components of the stimulus and is inversely related to its ability to elicit the illusion. This gives the model the ability to describe individuals tested with different stimuli, even if the physical properties of the stimulus are unknown.

The NED model's use of three simple parameters—sensory noise, disparity threshold, and stimulus disparity—highlights the differences between it and other models of the McGurk illusion, such as the seminal model of auditory and visual speech known as the fuzzy logical model of perception (FLMP; Massaro, 1998). The FLMP model creates distinct parameters for the auditory and visual component for each stimulus for each participant, resulting in a very large number of parameters (thirteen times more parameters than the NED model for the data described in this paper). Because each FLMP parameter represents the interaction between a particular stimulus and a particular individual, there is no way to use the parameters to separately examine individual differences and differences between stimuli.

We tested the NED model against real data and measured whether independent stimulus and participant parameters could explain a significant portion of the variance in McGurk perception. We demonstrate the utility of the model for group comparisons using published data on a clinical population, individuals with autism spectrum disorder.

Methods

Three types of model parameters

The noisy encoding of disparity (NED) model of the McGurk Illusion contains three types of parameters. The first parameter, the disparity of each stimulus (*D*), captures the likelihood that the auditory and visual components of the syllable produce the McGurk effect, averaged across all presentations of the stimulus to all participants.

The second parameter describes the sensory noise in each participant (σ). When an individual perceives multisensory speech, the auditory and visual features are measured with noise (Bejjanki, Clayards, Knill, & Aslin, 2011; Ma et al., 2009), resulting in variability in the measured stimulus disparity (Figure 1A). Across many trials, the distribution of measured strengths will be Gaussian in shape, centered at the true stimulus disparity, D_i (*i* indexes the stimuli), with standard deviation equal to the individual sensory noise, σ_j (*j* indexes the participants). For any participant, the amount of sensory noise is assumed to be constant across stimuli.

The third parameter, the disparity threshold (*T*), describes each participant's prior probability for fusing the auditory and visual components of the syllable by a fixed threshold placed along the stimulus disparity axis (Figure 1B). If the measured disparity is below this

threshold, the auditory and visual speech cues are fused, and the individual reports a fusion percept. To predict proportions of fusion perception, we calculate the probability that the measured disparity for a given stimulus is below a participant's threshold (T_j; Figure 1C):

$$p\left(x<T_j|D_i\right)=\int_{-\infty}^{T_j} N\left(x;D_i,\sigma_j\right) dx$$

where N is the Normal (Gaussian) distribution with mean D_i and standard deviation σ_j. The invariance of the disparity threshold and sensory noise across stimuli allows the model to predict a participant's fusion proportion for any stimulus with known strength, even if the participant has not seen the stimulus. Because the stimulus disparities are fixed across participants, they cannot fit participant variability; stimulus disparities and participant disparity thresholds are independent.

Model fitting—All model fitting was done in R (R Core Team, 2012); source code for the NED model is freely available at http://openwetware.org/wiki/Beauchamp:NED. Fits were obtained by minimizing the squared error between the model predictions and the behavioral data (proportion of fusion responses for each stimulus across all trials). This is similar to fitting by maximizing the log likelihood of the model, except in the cases in which the model predicts a fusion proportion of 0.0 or 1.0; in which case the log likelihood may go to negative infinity, but the squared error remains finite. We restrict the range of the parameters so that the participant parameters do not go to infinity when participants have 0.0 or 1.0 mean fusion proportion (stimulus disparities and individual thresholds: 0 to 2; sensory noise: 0 to 1). If parameters were allowed to go to infinity, it would not change the overall model fit but would make it impossible to calculate the mean or standard deviation of parameters, a necessity for inter-participant or inter-group comparisons.

The values for all three parameter types have the same units (disparity), but the scale of these units is arbitrary. The scale was fixed during the fitting process by setting initial values for the stimulus disparities. For each stimulus we exponentiated (base e) 1 minus the mean fusion proportion for each stimulus, and then subtracted 1. This transform ensures non-negative disparity values, mapping 1.0 fusion proportions to 0 disparity and 0.0 fusion proportions to 1.72. Next, the participant parameters were estimated, followed by fitting each of the stimulus disparities. This was followed by 9 cycles of fitting the participant parameters and then the stimulus parameters to converge on the best-fitting parameters. To guard against fitting to local optima, we created 48 initial guesses for the initial vector of stimulus disparities using a sample from a Gaussian distribution with standard deviation 0.05 and mean equal to the transformed stimulus fusion proportions.

Parameter testing and generalization—In order to estimate the importance of the three types of parameters (D, T, and σ) we calculated the fit error between the full NED model and three different model variants, each of which held one of the parameter types fixed. Additionally, a cross-validation approach was used to test the generalization of the individual-level parameters to untrained stimuli. During cross validation, each model was fit to a subset of the full dataset (the training set) and the fitted models were used to predict the fusion proportion on the remaining data (the testing set). The training set excluded a single

randomly-selected stimulus from each participant in the full dataset and the testing set was the fusion proportion for this held-out stimulus. We calculated the error between the model's prediction for the held-out fusion proportion and the overall mean fusion proportion for each participant. To estimate average performance, we completed 25 cross-validation runs.

Data description—We used data from a large-scale laboratory study of college-aged adults that measured McGurk illusion perception (Basu Mallick, Magnotti, & Beauchamp, *submitted*). Participants completed 10 trials with each McGurk stimulus, randomly intermixed with trials containing congruent auditory-visual speech and incongruent (but non-McGurk eliciting) auditory-visual speech. $N = 66$ participants were tested with 14 McGurk stimuli, $N = 77$ were tested with 9 McGurk stimuli and $N = 22$ were tested with 10 McGurk stimuli. To fit the model, we treated the untested stimuli for each participant as missing data.

Results

There was a great deal of variability in the behavioral data, providing a challenge to a model that must use identical stimulus parameters for all individuals and identical individual parameters for all stimuli. As shown in Figure 2A, there was a large range of fusion proportions for different stimuli, from 0.17 to 0.81. Within each single stimulus, there was a high degree of variability across individuals, with McGurk perception varying 40% from the mean on average (mean SD = 0.39). This variability across participants is illustrated in Figure 2B, showing that participants' mean fusion proportions across stimuli ranged from the lowest possible value (0.0, no fusions) to the highest possible value (1.0, 100% fusion). Despite these challenges, the model provided an overall good fit to the behavioral data (average root mean square error across stimuli, RMSE = 0.026; across participants, RMSE = 0.032).

The NED model makes two related but distinct claims: that individual participant effects are consistent across stimuli, and that stimulus effects are consistent across participants. We first examined if participant parameters are consistent across stimuli *i.e.* if participant 1 has more fusion than participant 2 for stimulus A, then participant 1 should also have more fusion than participant 2 for stimulus B. We calculated each participant's rank (out of 165) for each stimulus and then compared it to that participant's overall rank (averaged across stimuli). There was a significant positive correlation between the participant ranks at each stimulus and across all stimuli (mean Spearman correlation 0.65 ± 0.04 SEM; bootstrap mean = 0.26; bootstrap *p*-value = 10^{-137}). Next, we examined the assumption that stimulus effects are consistent across participants *i.e.* if stimulus A is weaker than stimulus B in participant 1, it should also be weaker in participant 2. We calculated each stimulus's rank (out of 14) for each participant and then compared it to that stimulus's overall rank (averaged across participants). There was a significant positive correlation between the stimulus ranks for each participant and across all participants (mean Spearman correlation 0.64 ± 0.02; bootstrap mean = 0.07; bootstrap *p*-value = 10^{-84}).

Parameter assessment

To assess the importance of the three parameters to the observed fit, we created three model variants, each holding a single parameter type fixed at single value (akin to testing full and reduced regression models). We calculated the percent reduction in fit error between the full NED model and the model variant with the variable held fixed to measure the importance of that variable (Figure 2C). Individual stimulus disparities were the most important (47% reduction), then individual disparity thresholds (46% error reduction), and finally individual sensory noise (25% reduction). Although sensory noise was less important than the other parameters for reducing prediction error, it was critical for realistic model predictions. Without the sensory noise parameter, the model would predict all-or-none fusion perception (0.0 or 1.0) for a given individual viewing a particular stimulus, but only 12 out of 165 participants (7%) demonstrated all-or-none responding across stimuli. Paired *t*-tests confirmed that fixing any of the parameter types resulted in a significantly more fit error [all $ts(164) > 13.6$, $ps < 10^{-15}$] across participants.

The sensory noise parameter of the model does not correspond directly to the noise along a single physical dimension, but rather represents the combination of noise sources from multiple independent sources (e.g., auditory encoding, visual encoding, varying attention). Figure 3 shows that the sensory noise parameter is related to the variability (mean binomial standard deviation across stimuli) in fusion perception (Spearman's $r = 0.59$, $p = 10^{-15}$), but not to the average fusion proportion (Spearman's $r = 0.11$, $p = 0.15$). This dissociation suggests that individuals may differ not only on disparity threshold (related to mean fusion proportion) but also in the variability of their fusion proportion.

Predicting novel stimuli

One important advantage of the NED model is that it allows prediction of unseen stimuli; this can be important in behavioral training paradigms in which generalization is tested with new stimuli. The model was fit to the behavioral data with a random stimulus left out for each participant. Then, these fitted parameters were used to predict the fusion proportion of the left-out stimulus. Across participants the model explained 53% of the variability in fusion scores on the held-out stimuli (mean $R^2 = 0.53 \pm 0.01$) and individual predictions for novel stimuli were within 20% on average (mean absolute error = 0.19 ± 0.01).

Fitted parameter values

Next, we examined the three types of parameters used by the NED model. Although the absolute value of the parameters is not important because the units are arbitrary, their relative values across stimuli and participants are of interest for comparing individuals or groups. Stimulus disparity ranged from 0.26 for the strongest stimulus to 1.25 for the weakest stimulus with mean of 0.81. The disparity threshold parameter ranged from 0.0 for participants who never perceived the illusion to 2.0 for participants who always perceived the illusion, with a mean disparity threshold across participants of 0.72, below the mean stimulus disparity. The sensory noise parameter values ranged from 0.001 for participants with the lowest sensory noise to 1.0 for participants with the highest sensory noise. The distribution of sensory noise across participants was positively skewed (median = 0.12; mean = 0.23).

Model application to group data

There are a number of conflicting results in the McGurk effect literature. For instance, several groups have reported *less* McGurk perception in individuals with autism spectrum disorder (ASD) compared to matched controls (Bebko, Schroeder, & Weiss, 2014; Irwin et al., 2011; Mongillo et al., 2008; Stevenson, Siemann, Schneider, et al., 2014; Stevenson, Siemann, Woynaroski, et al., 2014), while others have reported similar or *more* McGurk perception in individuals with ASD (Saalasti et al., 2012; Taylor, Isaac, & Milne, 2010; Woynaroski et al., 2013). Differences in experimental methods or statistical power might change the reported strength of an effect, but different directions of an effect are much more difficult to explain.

The NED model can explain group differences using the disparity threshold or sensory noise parameters with different results. A group difference in disparity threshold will result in changes in the proportion of McGurk perception that are in the same direction for any stimulus. A group difference in sensory noise will cause different effects for different stimuli. For stimuli with low disparity, higher noise will cause less McGurk perception (as noise drives the perceived disparity above threshold on some trials) while for stimuli with high disparity, higher noise will cause more McGurk perception (as noise drives the perceived disparity below threshold on some trials).

To apply the model, we selected two studies with seemingly contradictory results. Irwin et al. (2011) reported *lower* fusion proportions for ASD than TD (0.56 *vs.* 0.88, $p = 0.01$) while Woynaroski et al. (2013) reported *higher* fusion proportions for ASD and TD participants (0.38 *vs.* 0.28, $p = 0.37$). We assumed that both studies were representative of the larger ASD and typically developing (TD) populations and created two hypothetical participants (one ASD and one TD), assigning them each fusion proportions equal to the mean of their group.

Fit to these data (Figure 4), the model estimated similar disparity thresholds for the ASD and TD participants (difference of 0.1) but much larger sensory noise for the ASD participant (4-fold greater). For the stimulus parameter, the model estimated a higher disparity value for the Woynaroski et al. stimulus (Figure 4A) than for the Irwin et al. stimulus (Figure 4B). The Woynaroski et al. stimulus is above threshold for both TD and ASD, but because the ASD participant has high sensory noise, many of the perceived disparities fall below threshold, resulting in a higher fusion proportion for ASD. The Irwin et al. stimulus is below threshold for both participants, but because of the ASD participant's high sensory noise, many of the perceived disparities fall above threshold, resulting in a lower fusion proportion.

Discussion

We describe the noisy encoding of disparity (NED) model of McGurk fusion perception. The NED model captures both within-individual and between-individual variability using three types of parameters: two for each participant (disparity threshold and sensory noise) and one for each stimulus (disparity). The model was able to predict fusion perception on novel stimuli (not fit by the model for that individual) within 20% of their actual value.

The two key assumptions of the model are that individuals have characteristics (captured by the disparity threshold and sensory noise model parameters) that are independent of the particular stimulus they are viewing, and that McGurk stimuli have intrinsic properties that are independent of the individual viewing them (captured by the disparity model parameter). These assumptions allow NED to predict fusion perception for stimuli that an individual has never seen. For example, it predicts that a participant with a high threshold will be likely to perceive the McGurk effect when shown a stimulus with low disparity.

Explanation of model parameters

The parameters used by the NED model are not response probabilities. Instead, they are the product of an analysis of different factors that may contribute to McGurk perception. The first participant parameter, amount of sensory noise, explains a notable aspect of the McGurk illusion: individuals do not always report the same percept when presented with the same stimulus. We account for this phenomenon using sensory noise because it is well recognized that the brain encodes external stimuli in a noisy fashion (Angelaki et al., 2009; Knill & Pouget, 2004; Seilheimer, Rosenberg, & Angelaki, 2014). For a stimulus that evokes average illusion perception, large values of sensory noise correspond to highly variable McGurk perception across trials. The second participant parameter, the disparity threshold, explains another notable aspect of the illusion: that some *individuals* rarely perceive the illusion while others often perceive it. The stimulus disparity parameter is necessary to explain the observation that some *stimuli* rarely evoke the illusion while others often evoke it.

By separately modeling inter-participant differences and inter-stimulus differences, NED can explain seemingly contradictory results in the literature on McGurk perception in individuals with ASD. The model fits suggest that ASD individuals have greater sensory noise, resulting in less McGurk perception when tested with stimuli with low disparity and more McGurk perception when tested with stimuli with high disparity. This suggestion is supported by evidence from recent functional imaging and anatomical studies (Dinstein et al., 2012; Geschwind & Levitt, 2007).

Utility of a simplified model of the McGurk Effect

One motivation for developing a simplified model of the McGurk effect is that existing models do not allow generalization (*i.e.* predicting what an individual will perceive before they have ever seen a given stimulus). To allow this, we collapse all stimulus differences into a single parameter, disparity. This contrasts with physical, bottom-up approaches (Jiang & Bernstein, 2011; Omata & Mogi, 2008) that attempt to characterize the many different physical dimensions across which stimuli differ. In principle, it should be possible to relate our single disparity parameter to the audiovisual properties of each stimulus.

The simplification of the McGurk illusion to a single dimension allows us to relate illusion perception across stimuli with diverse properties (e.g., syllables used, number of syllables in each stimulus, different talkers). Parameter values will reflect individual differences along this new dimension, rather than being tied to a physical measurement of the auditory or visual stimulus. In this way, the procedure is akin to other dimension-reduction techniques

(e.g., principal components analysis) that project high-dimensional data onto a lower-dimensional space.

It is interesting to consider how common stimulus manipulations, like adding stimulus noise, would be handled by the NED model. Intuitively, it might seem that adding sensory noise to a multisensory stimulus should not change the perceived disparity, and thus the model parameters would remain unchanged. However, Bayes-optimal integration models predict that increasing noise actually *increases* individuals' perception of a common cause (Magnotti, Ma, & Beauchamp, 2013; Shams & Beierholm, 2010), as well as changing the weight assigned to each cue. The model would therefore assign a different disparity rating to clear and noisy versions of a given stimulus. Because the model does not consider the cue weighting that occurs during optimal integration directly, it does not address the differential effect of adding visual vs. auditory noise: adding auditory noise increases McGurk illusion (by lowering the relative weight to the auditory cue), adding visual noise decreases McGurk illusion (by increasing the relative weight to the auditory cue). A two-step model that incorporates both causal inference and optimal integration is necessary to capture these effects (Kording et al., 2007; Nahorna, Berthommier, & Schwartz, 2012). For typical McGurk studies that neither collect unisensory recognition data nor manipulate sensory reliability, however, such a model could not be fit.

Comparison with other models

The most popular existing models of multisensory speech perception, the fuzzy logical model of perception (FLMP; Massaro, 1998) and its variants (Schwartz, 2010), fit each individual's response to each stimulus independently. In addition to eliminating the ability to predict responses to new stimuli, the matrix of generated parameters lacks a straightforward interpretation. For example, with the current dataset, a comparable FLMP-type model would require at least 28 parameters *per participant* (one for the auditory and visual weights for each stimulus) resulting in a total of 4620 parameters. NED requires thirteen-fold fewer parameters (344). To fit this large number of parameters, FLMP-type models require collecting unisensory data for each participant/stimulus pair. For the current dataset with 14 McGurk stimuli, FLMP-type models would require collecting data from an additional 28 conditions, tripling data acquisition time, a potentially insurmountable problem for certain populations (*e.g.*, young children, clinical populations).

The two subject parameters in the NED model are immediately relatable to studies of the neural underpinnings of the McGurk effect. For instance, Nath and Beauchamp (2012) showed that individuals with lower McGurk fusion proportions had correspondingly lower activity in the left STS, an area critical for multisensory integration during speech perception. The response amplitude from this region could provide a neural correlate of the disparity threshold parameter in our model. Using neuroimaging, it is also possible to measure trial-to-trial variability in the response to identical stimuli (Dinstein et al., 2012; Keil et al., 2012). This could provide a neural correlate of the sensory noise parameter in our model.

Relationship to item response theory

The NED model bears some resemblance to one-dimensional item-response theory models (IRT; Gelman & Hill, 2007) that characterize individual ability and item difficulty. Applied to the McGurk illusion, an IRT model would characterize each individual with a susceptibility level and each stimulus with an efficacy (how strongly it elicits the illusion). This IRT model would predict that two participants with the same susceptibility level would have exactly the same amount of McGurk perception for a given stimulus. Our data suggests that this is not the case, and that including a sensory noise parameter (which allows perception to vary between individuals with the same disparity threshold) accounts for a significant amount of variance (Figure 2C).

Conclusion

Previous studies describing large variability of fusion perception across stimuli and individuals illustrate the peril of comparing raw amount of McGurk perception between individuals or groups: combining raw fusion proportions from different stimuli confounds stimulus differences and individual/group differences. The NED model provides a way to extract independent parameters for participants (threshold and sensory noise) and stimuli (stimulus disparity). This independence allows for individual and group-level comparisons using measures related to individual differences rather than the properties of individual stimuli.

Acknowledgments

References

Angelaki DE, Gu Y, DeAngelis GC. Multisensory integration: psychophysics, neurophysiology, and computation. Current Opinion in Neurobiology. 2009; 19(4):452–458. doi: 10.1016/j.conb. 2009.06.008. [PubMed: 19616425]

Beauchamp MS, Nath AR, Pasalar S. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. Journal of Neuroscience. 2010; 30(7):2414–2417. doi: 30/7/2414 [pii] 10.1523/JNEUROSCI.4865-09.2010. [PubMed: 20164324]

Bebko JM, Schroeder JH, Weiss JA. The McGurk effect in children with autism and Asperger syndrome. Autism Res. 2014; 7(1):50–59. doi: 10.1002/aur.1343. [PubMed: 24136870]

Bejjanki VR, Clayards M, Knill DC, Aslin RN. Cue integration in categorical tasks: insights from audio-visual speech perception. PLoS One. 2011; 6(5):e19812. doi: 10.1371/journal.pone.0019812. [PubMed: 21637344]

Dinstein I, Heeger DJ, Lorenzi L, Minshew NJ, Malach R, Behrmann M. Unreliable evoked responses in autism. Neuron. 2012; 75(6):981–991. doi: 10.1016/j.neuron.2012.07.026. [PubMed: 22998867]

Gelman, Andrew; Hill, Jennifer. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press; Cambridge; New York:: 2007.

Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. Current Opinion in Neurobiology. 2007; 17(1):103–111. doi: 10.1016/j.conb.2007.01.009. [PubMed: 17275283]

Green KP, Kuhl PK, Meltzoff AN, Stevens EB. Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect. Perception and Psychophysics. 1991; 50(6):524–536. [PubMed: 1780200]

Irwin JR, Tornatore LA, Brancazio L, Whalen DH. Can children with autism spectrum disorders "hear" a speaking face? Child Dev. 2011; 82(5):1397–1403. doi: 10.1111/j. 1467-8624.2011.01619.x. [PubMed: 21790542]

Jiang J, Bernstein LE. Psychophysics of the McGurk and other audiovisual speech integration effects. Journal of Experimental Psychology: Human Perception and Performance. 2011; 37(4):1193–1209. doi: 10.1037/a0023100. [PubMed: 21574741]

Keil J, Muller N, Ihssen N, Weisz N. On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. Cerebral Cortex. 2012; 22(1):221–231. doi: 10.1093/cercor/bhr125. [PubMed: 21625011]

Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neurosciences. 2004; 27(12):712–719. doi: 10.1016/j.tins.2004.10.007. [PubMed: 15541511]

Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. PLoS One. 2007; 2(9):e943. doi: 10.1371/journal.pone.0000943. [PubMed: 17895984]

Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. PLoS One. 2009; 4(3):e4638. doi: 10.1371/journal.pone.0004638. [PubMed: 19259259]

MacDonald J, McGurk H. Visual influences on speech perception processes. Perception and Psychophysics. 1978; 24(3):253–257. [PubMed: 704285]

Magnotti, John F.; Ma, Wei Ji; Beauchamp, Michael S. Causal inference of asynchronous audiovisual speech. Frontiers in Psychology. 2013; 4:798. [PubMed: 24294207]

Mainen ZF, Sejnowski TJ. Reliability of spike timing in neocortical neurons. Science. 1995; 268(5216):1503–1506. [PubMed: 7770778]

Massaro, DW. Perceiving talking faces: from speech perception to a behavioral principle. MIT Press; Cambridge, Mass.: 1998.

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264(5588):746–748. [PubMed: 1012311]

Mongillo EA, Irwin JR, Whalen DH, Klaiman C, Carter AS, Schultz RT. Audiovisual processing in children with and without autism spectrum disorders. J Autism Dev Disord. 2008; 38(7):1349–1358. doi: 10.1007/s10803-007-0521-y. [PubMed: 18307027]

Munhall KG, Gribble P, Sacco L, Ward M. Temporal constraints on the McGurk effect. Perception & psychophysics. 1996; 58(3):351–362. [PubMed: 8935896]

Nahorna O, Berthommier F, Schwartz JL. Binding and unbinding the auditory and visual streams in the McGurk effect. J Acoust Soc Am. 2012; 132(2):1061–1077. doi: 10.1121/1.4728187. [PubMed: 22894226]

Nath AR, Beauchamp MS. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. NeuroImage. 2012; 59(1):781–787. doi: 10.1016/j.neuroimage. 2011.07.024. [PubMed: 21787869]

Omata, Kei; Mogi, Ken. Fusion and combination in audio-visual integration. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science. 2008; 464(2090):319–340.

R Core Team. R Foundation for Statistical Computing; Vienna, Austria: 2012. R: A language and environment for statistical computing.

Saalasti S, Katsyri J, Tiippana K, Laine-Hernandez M, von Wendt L, Sams M. Audiovisual speech perception and eye gaze behavior of adults with asperger syndrome. J Autism Dev Disord. 2012; 42(8):1606–1615. doi: 10.1007/s10803-011-1400-0. [PubMed: 22068821]

Schwartz JL. A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. J Acoust Soc Am. 2010; 127(3):1584–1594. doi: 10.1121/1.3293001. [PubMed: 20329858]

Seilheimer, Robert L.; Rosenberg, Ari; Angelaki, Dora E. Models and processes of multisensory cue combination. Current Opinion in Neurobiology. 2014; 25:38–46. [PubMed: 24709599]

Sekiyama K. Cultural and linguistic factors in audiovisual speech processing: the McGurk effect in Chinese subjects. Perception and Psychophysics. 1997; 59(1):73–80. [PubMed: 9038409]

Shams L, Beierholm UR. Causal inference in perception. Trends Cogn Sci. 2010; 14(9):425–432. doi: S1364-6613(10)00147-6 [pii] 10.1016/j.tics.2010.07.001. [PubMed: 20705502]

Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT. Multisensory temporal integration in autism spectrum disorders. Journal of Neuroscience. 2014; 34(3):691–697. doi: 10.1523/JNEUROSCI.3615-13.2014. [PubMed: 24431427]

Stevenson RA, Siemann JK, Woynaroski TG, Schneider BC, Eberly HE, Camarata SM, Wallace MT. Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. J Autism Dev Disord. 2014; 44(6):1470–1477. doi: 10.1007/s10803-013-1992-7. [PubMed: 24218241]

Stevenson RA, Zemtsov RK, Wallace MT. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. Journal of Experimental Psychology: Human Perception and Performance. 2012; 38(6):1517–1529. doi: 10.1037/a0027339 2012-05374-001 [pii]. [PubMed: 22390292]

Taylor N, Isaac C, Milne E. A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. J Autism Dev Disord. 2010; 40(11):1403–1411. doi: 10.1007/s10803-010-1000-4. [PubMed: 20354776]

Tremblay C, Champoux F, Voss P, Bacon BA, Lepore F, Theoret H. Speech and non-speech audio-visual illusions: a developmental study. PLoS One. 2007; 2(1):e742. doi: 10.1371/journal.pone. 0000742. [PubMed: 17710142]

Woynaroski TG, Kwakye LD, Foss-Feig JH, Stevenson RA, Stone WL, Wallace MT. Multisensory Speech Perception in Children with Autism Spectrum Disorders. J Autism Dev Disord. 2013 doi: 10.1007/s10803-013-1836-5.
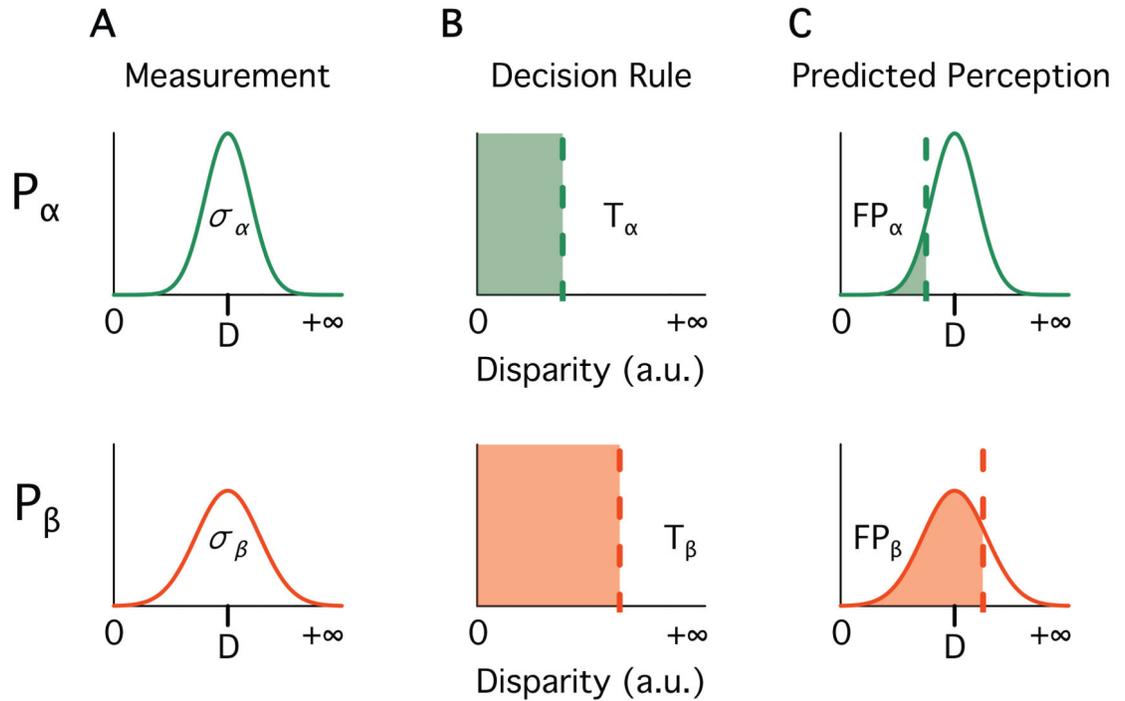
Figure 1.
The noisy encoding of disparity (NED) model explains proportion of McGurk perception with three parameters, shown for two hypothetical participants (P$_\alpha$, top row, green color; P$_\beta$, bottom row, red color). All variables are defined in arbitrary units of disparity. **A**. The first parameter is the disparity of each stimulus. Tick mark on the *x*-axis shows the disparity of hypothetical stimulus S (identical across participants). The second parameter is the sensory noise (σ, standard deviation of the curve) of each participant. The measured stimulus disparity is a noisy approximation of the true stimulus disparity. **B**. The third parameter is the disparity threshold for each participant (T; vertical dashed lines). Participants report a fused percept whenever the measured stimulus disparity is below this threshold. **C**. Predicted McGurk fusion proportion (FP, shaded region) calculated by combining **A** and **B**.
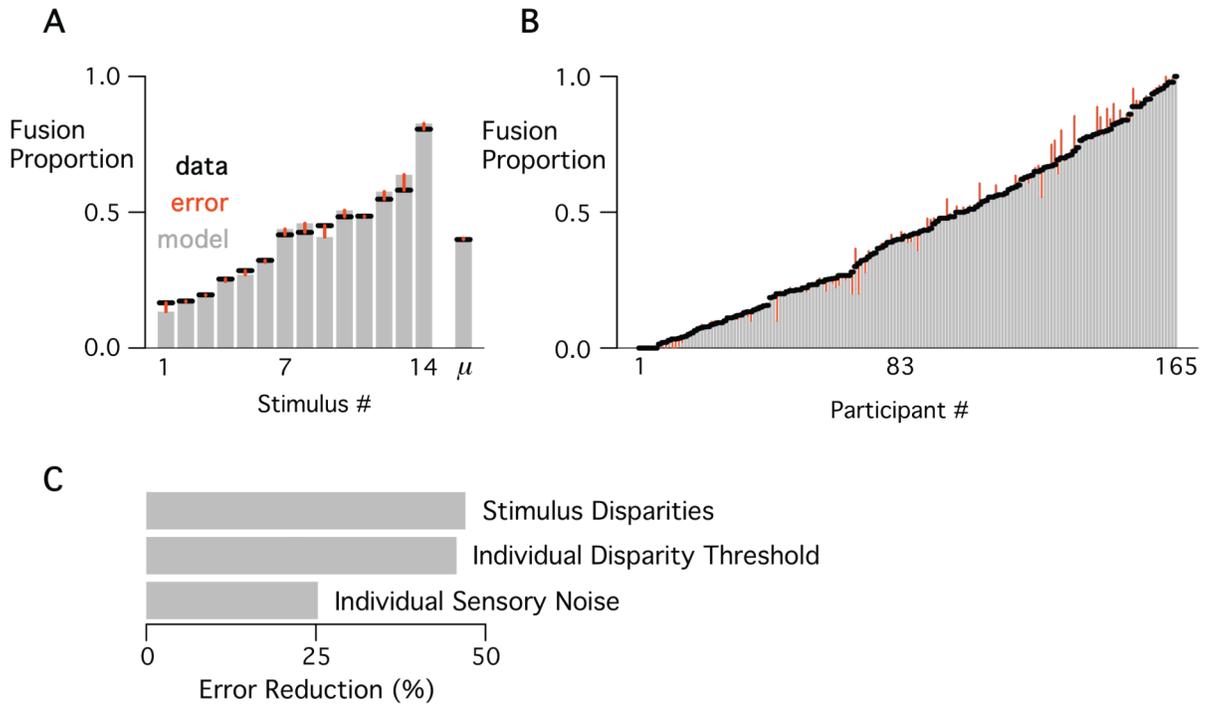
Figure 2.
The NED model fit to real behavioral data. **A.** Mean fusion proportion (black lines) and mean model predictions (gray bars) across participants for each stimulus and the mean across all participants and stimuli (μ). Red lines show prediction error (across stimuli, mean RMSE = 0.03; overall mean RMSE = 0.001). **B.** Mean fusion proportion across stimuli, with model prediction and prediction error plotted for every participant (across participants, mean RMSE = 0.03). Participants ordered by mean fusion proportion. **C**. The percent reduction of total RMSE when a parameter was allowed to vary *vs.* when it was held fixed (across all participants).
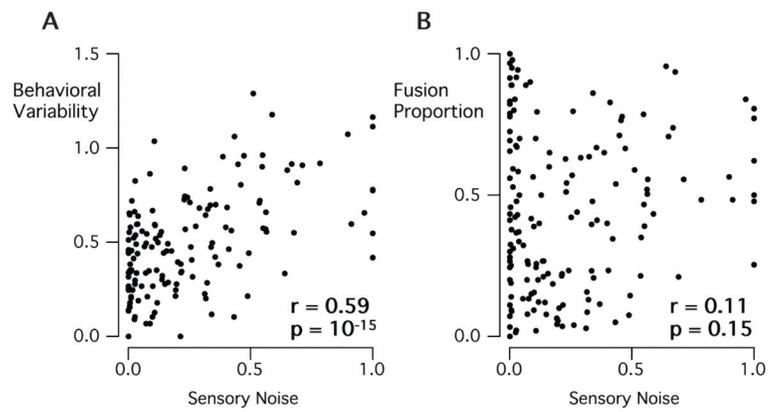
Figure 3.
Relationship between sensory noise and McGurk fusion perception. **A.** Sensory noise is significantly correlated (Spearman correlation) with behavioral variability (mean binomial standard deviation across stimuli) across participants. **B.** Sensory noise is only weakly correlated to mean fusion proportion across participants.
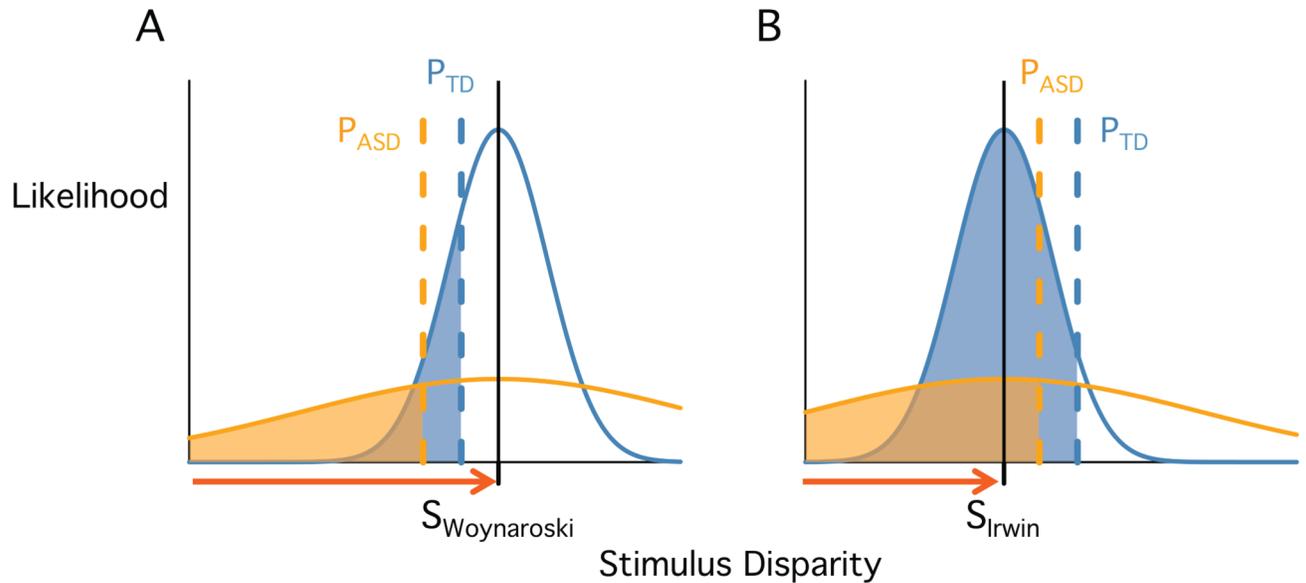
Figure 4.
The NED model can explain seemingly contradictory results in studies using different stimuli. **A**. In the study of Woynaroski et al., typically-developing (TD) participants had a *lower* mean McGurk fusion proportion than participants with autism spectrum disorder (ASD). Two representative participants (one TD, blue; one with ASD, orange) were assigned the published mean fusion proportion and fit with the model. The model estimated the true stimulus disparity (black tick mark and vertical line) the threshold for each participant (dashed lines) and each participant's sensory noise (width of curves, see Figure 1). The stimulus disparity is above threshold for both participants, but the greater sensory noise for ASD leads to a *higher* proportion of trials with measured disparities below threshold (orange shaded region larger than blue shaded region). **B**. In the study of Irwin et al., TD participants had a higher mean McGurk fusion rate than participants with ASD. The model is fit simultaneously to the data in (**A**) and (**B**), incorporating the assumption that participant thresholds and sensory noise are equal across studies. The fitted stimulus disparity of S_{Irwin} is lower than $S_{Woynaroski}$ (red arrows). The disparity of S_{Irwin} is below the disparity threshold for both participants, but the greater sensory noise for the ASD participant leads to a *lower* proportion of trials with measured disparities below threshold (orange shaded region smaller than blue shaded region).