# Spontaneous Otoacoustic Emissions Reveal an Efficient Auditory Efferent Network

Viorica Marian*[1], Tuan Q. Lam[2], Sayuri Hayakawa[1], & Sumitrajit Dhar[1]

[1]Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL
[2]Department of Psychological Sciences, Loyola University, New Orleans, LA

*Correspondence regarding this manuscript should be addressed to Viorica Marian, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208, United States of America. Phone: (847)-491-2420.
E-mail: v-marian@northwestern.edu

# Abstract

**Purpose.** Understanding speech often involves processing input from multiple modalities. The availability of visual information may make auditory input less critical for comprehension. This study examines whether the auditory system is sensitive to the presence of complementary sources of input when exerting top-down control over the amplification of speech stimuli.

**Method.** Auditory gain in the cochlea was assessed by monitoring spontaneous otoacoustic emissions (SOAEs), which are by-products of the amplification process. SOAEs were recorded while thirty-two participants (23 female; $M_{age}$=21.13) identified speech sounds such as "ba" and "ga." The speech sounds were presented either alone or with complementary visual input, as well as in quiet or with six-talker babble.

**Results.** Analyses revealed that there was a greater reduction in the amplification of noisy auditory stimuli compared to quiet. This reduced amplification may aid in the perception of speech by improving the signal-to-noise ratio. Critically, there was a greater reduction in amplification when speech sounds were presented bimodally with visual information relative to when they were presented unimodally. This effect was evidenced by greater changes in SOAE levels from baseline to stimuli presentation in Audio-Visual trials relative to Audio-Only trials.

**Conclusions.** The results suggest that even the earliest stages of speech comprehension are modulated by top-down influences, resulting in changes to SOAEs depending on the presence of bimodal or unimodal input. Neural processes responsible for changes in cochlear function are sensitive to redundancy across auditory and visual input channels, and coordinate activity to maximize efficiency in the auditory periphery.

# Introduction

Sensation is typically thought of as a bottom-up process by which sensory organs convert physical inputs into neural signals that are then interpreted by the brain. However, high-level cognitive processes such as attention can have profound effects on how we process sensory inputs (Honrubia and Elliott, 1968). For example, focusing attention on a visual task can reduce sensitivity to irrelevant auditory stimuli (Puel, Bonfils & Pujol, 1988). What is not known is whether the auditory system is sensitive to relevant, but redundant cross-modal information, such as when processing audio-visual speech. Put simply, do we turn down the volume of auditory input when visual aids are present?

To investigate this question, we examined whether different sensory conditions lead to changes in sounds emitted spontaneously by the ear, called spontaneous otoacoustic emissions (SOAE; Zhao & Dhar, 2011; Zhao & Dhar, 2012). The outer hair cells (OHC) in the inner ear provide amplification to auditory input before they are converted to neural signals. Otoacoustic emissions (OAE) are a by-product of this amplification process and can be recorded in the ear canal using a sensitive microphone (Kemp, 1978). The brain can influence OHC-mediated amplification, and therefore OAEs, through the action of the auditory efferent network which originates in the cortex and terminates in the cochlea. Fibers of the cholinergic medial olivocochlear (MOC) bundle terminate directly on OHC and inhibit amplification and OAEs (Guinan Jr., 2006). Important for the purposes of this report, OAEs have become the *de facto* tool of choice in evaluating the auditory efferents in humans, allowing a non-invasive and reliable examination of central control over peripheral amplification (Deeter, Abel, Calandruccio & Dhar, 2009; Zhao & Dhar, 2011; Zhao & Dhar, 2012). We chose to specifically investigate spontaneous otoacoustic emissions over other classes of otoacoustic emissions (e.g., transient evoked otoacoustic emissions, distortion-product otoacoustic emissions) because they are highly sensitive to MOC modulation (Harrison & Burns, 1993) and require no external stimulation of the monitored ear (Kemp, 1978). Acoustic stimulation of the ear would have altered the internal signal to noise ratio for the speech-in-noise task. Monitoring spontaneous otoacoustic emissions afforded the opportunity to monitor changes in cochlear gain without adding any additional external stimulation. Furthermore, the effect of visual input on SOAEs is less established compared to other types of emissions (Meric and Collet, 1994).

Various functional roles of the auditory efferents have been demonstrated in laboratory animals, such as protecting the inner ear against the toxic effects of noise (Maison and Liberman, 2000) and preventing premature aging of the inner ear (Liberman, Liberman, & Maison, 2014). In humans, some reports suggest that the efferents assist in signal (speech) detection in noise (Kawase, Delgutte, & Liberman, 1993; Kumar and Vanaja, 2004; but see Wagner, Frey, Heppelmann, Plontke, & Zenner, 2008). Detecting a transient signal such as speech in a noisy environment can become difficult if the output of the auditory nerve is saturated by the noise (Guinan Jr., 2006). Efferent inhibition of the sensory response to noise can restore the ability to detect transient signals (Kawase et. al, 1993). Here, we investigate the possible role of the efferent network in comprehending auditory speech stimuli when it is paired with visual speech stimuli.

While there is evidence that attention to visual input can lead to a reduction in cochlear gain in humans, previous experiments utilized auditory stimuli that are functionally or ecologically irrelevant (e.g. listening to clicks while looking at letters on a screen; Meric and Collet, 1994). In real life, auditory and visual input often provide *complementary* information, and may thus provide less of a reason to reduce cochlear gain when processing visual information. For example, when processing speech in noise, studies have shown a clear benefit of audio-visual (AV) speech in comparison to audio-only (AO) speech (Grant and Seitz, 2000; Sumby and Pollack, 1954) suggesting that the two inputs are not fully redundant. In fact, when AV speech is available, listeners often integrate both sources so that if the auditory and visual information are mismatched, listeners sometimes perceive an intermediate sound (McGurk and MacDonald, 1976). However, auditory and visual speech do contain some redundant information, and with greater redundancy, the benefit for AV speech is weakened (Grant and Walden, 1996; Grant, Walden & Seitz, 1998). As such, efficient attenuation of cochlear gain may be possible without any consequence to communication effectiveness. If, as predicted, the efferent system inhibits the amplification of redundant information, we ought to observe reduced SOAEs in response to audio-visual stimuli as compared to audio-only stimuli.

## Methods

### Ethical Approval

The experiment was carried out with the approval of Northwestern University's Institutional Review Board following approved guidelines and regulations. Informed consent was obtained from all participants prior to participation.

### Participants

Fifty-one adults participated in this study. Of the initially recruited participants, eight males and eleven females were excluded because they did not have SOAEs in baseline measurements. The remaining 32 participants (23 female) were between the ages of 18-26 (M =

21.13, SD = 2.14) and had normal hearing, as indicated by both self-reports and the presence of SOAEs, which are strong indicators of functional hearing.

## Materials

Stimuli consisted of audio-visual and audio-only speech syllables produced by a female native speaker of English. Six syllables: /ba/, /da/, /ga/, /pa/, /ta/, and /ka/ were presented. The consonant in each of these syllables is a stop consonant, which is produced by creating a complete stoppage of airflow in the vocal tract. The vowel /a/, which was used in all syllables, is a low back vowel and is acoustically the loudest vowel in English. Signal level was normalized at 60 dB peak to peak. Audio-only stimuli were obtained by extracting the audio from the videos. Instead of a video, participants were shown a static image of the speaker with the mouth closed while only the audio stimulus was played. This was done to match the two groups' experiences as closely as possible while making sure that only the audio-visual condition received visual information that would be useful in the comprehension of speech (Davis & Kim, 1999).

We manipulated two factors in this study: AV versus AO speech and quiet versus noise. For noise trials, the stimuli were presented in six-talker babble at 70 dB (+10 dB compared to the stimuli or -10 dB SNR). The babble noise began approximately 1500 milliseconds before the onset on the target syllable and continued for another 700 milliseconds after the end of the target syllable. For AV stimuli we presented the original video of the speaker producing a speech sound.

## Apparatus

Videos were displayed using a 27-inch iMac computer running MATLAB 2010. The screen resolution on the computer was set at 2560 x 1440 pixels. Sound was delivered unilaterally to a participant’s ear through an ear bud using an ER2 speaker made by Etymotic Research (Elk Grove Village, Illinois, USA). The medial olivocochlear response was continually recorded from the opposite ear canal using an ER-10B+ microphone (Etymotic Research, Elk Grove Village, Illinois, USA). Both the speaker delivering the speech sounds and the ER-10B+ probe were fitted to the subjects’ ear canals using comfortable and disposable foam tips. All recordings were conducted in a sound treated room.

## Procedure

Baseline SOAE records from each ear of each subject were established from three-minute recordings of the ear canal signal. A fast Fourier transform (FFT) using a 44100-point window (1 second) yielded a baseline SOAE spectrum with 1-Hz resolution from each ear (for an example, see Figure 1). For each subject, the ear with the highest number of SOAEs in the 1000-10000 Hz range was chosen to be monitored while speech sounds were presented to the opposite ear during the experiment (15 left ears and 17 right ears).

After baseline OAEs were obtained, participants began the speech perception task. At the start of a trial, participants first saw a motionless face. On noise trials, the onset of babble noise coincided with display onset and continued until 700 milliseconds after the end of the auditory stimuli. Exactly 1500 milliseconds after display onset, the face began producing a speech sound, and when finished, participants were presented with a six-item forced choice display from which participants had to indicate the sound they heard. After indicating their response, the next trial began.

There were a total of 240 trials that were split into ten blocks. After every block, participants were given a short break of approximately two minutes. At the halfway point of the experiment, participants were given a longer break (5-15 minutes) and allowed to move around and use the restroom.

## Data Analysis

For all OAE recordings, we first conducted a 22050-point short time Fourier transform on a window size of 16384 points and a hop size of 4096 points. We then identified the SOAEs that appeared between 1000-10000 Hz in the baseline condition in each participant. We computed the standard deviation across four 30-second periods in the baseline and removed the SOAEs with highly variable levels (SD > 6 dB), resulting in a total of sixty-eight SOAE frequencies. Of these, 37 were taken from the right ear (frequencies: 1068-6444 Hz) and 31 were taken from the left ear (frequencies: 1224-4538 Hz). Each participant had either one (N=9), two (N=13), three (N=7), or four (N=3) SOAE frequencies.

For each remaining SOAE, we computed the average level and frequency of the peak over the duration of the syllable. We also computed the average level and frequency of the peak over a 500-millisecond period before trial onset to use as the baseline for the trial. We did not analyze data for SOAEs whose peaks were within 3 Hz of a multiple of 60 Hz as these may have been generated by electrical noise. Analyses were conducted on the difference in peak level and frequency between syllable period and the baseline period. Level differences were computed in dB and changes in the frequency of the peak were computed in Hz. The analysis for decibel level differences was conducted using a multilevel regression model with a random intercept for individual SOAEs nested within subjects, as well as random slopes for noise and AV status. Multilevel regressions are particularly useful for analyzing data with clustered structures (Hox, 1998). In our case, this method allows us to account for multiple sources of variance coming from the fixed effects of noise and AV status, as well as the random effect of SOAE frequency, which was nested within subject. The model for frequency included all but the random slope for AV status, as the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) failed to converge. We additionally examined whether the number of SOAE frequencies per subject had an effect by first entering the number of SOAE frequencies as a predictor, and then rerunning the AV status and noise model on the residuals. In this way, we observed the effects of AV status and noise after controlling for the number of SOAE frequencies. Lastly, we examined whether the amount of SOAE inhibition (change in level from baseline to stimulus) affected accuracy in identifying the speech sound. This was done by entering accuracy as a binary dependent variable in a generalized linear model, with inhibition, noise, and AV status as fixed effects and SOAE frequency nested within subject as random effects. No random slopes were entered, as their inclusion precluded model convergence. We calculated effect sizes where appropriate using Judd, Westfall, and Kenny’s (2016) method of approximating Cohen's d, which accounts for variance from both fixed and random effects.

## Results

Overall, there was an upward shift in the SOAE frequency of 3.28 Hz (SE=0.09) and an average reduction in SOAE level of 3.41 dB (SE = 0.04) from baseline to when speech stimuli were presented in the opposite ear. The upward shift in frequency was larger for the noise condition than for the quiet condition ($\beta = 2.17$, SE = 0.06, t = 3.66, $p < 0.001$, d = .16). Similarly, there was a main effect of noise with greater inhibition in SOAE levels when the speech stimulus was presented with noise, versus in quiet ($\beta = -2.46$, SE = 0.36, t = -6.76, $p < 0.001$, d = -0.46). Table 1 displays the change in level (stimulus minus baseline), averaged across all SOAE frequencies in each of the four conditions. This finding is consistent with past research (Zhao and Dhar, 2011) and indicates that the auditory efferent system attenuates the amplification of noisy stimuli. Notably, we show that attenuation is not limited to tones and broadband noise (Kawase et. al, 1993) but also takes place with ecologically relevant speech sounds. The fact that such babble results in a reduction of cochlear gain underscores the influence of top-down processes for signal detection in noise generally, and speech perception in noise particularly.

Turning to the effect of complementary information, we find that as predicted, there was a small but noticeable inhibition of SOAE levels for the Audio-Visual conditions as compared to the Audio-Only conditions ($\beta = -0.17$, SE = 0.08, t = -2.13, $p = 0.041$, d = -0.04; Figure 2). Both the effects of noise and audiovisual status remain significant (both $p < .05$) after controlling for the number of SOAE frequencies per subject, and the number of frequencies had no significant effect on SOAE inhibition ($\beta = -0.05$, SE = 0.35, t = -0.16, $p = 0.874$). It should be noted that the audio-only effect captures the combined influence of an energetic activation of the auditory efferents at the level of the brainstem along with a more central attentional component. Thus, the magnitudes of the auditory and visual effects are not comparable prima facie. In general, comparative studies of auditory and visual attention have arrived at a diverse set of conclusions. For example, Walsh, Pasanen & McFadden (2015) found the magnitude of reduction in cochlear gain due to auditory and visual attention to be comparable. In contrast, Jedrzejczak et al. (2017) reported the effect of visual attention to be immeasurably small.

There was a significant effect of noise on accuracy, with better performance in quiet (98.1%) than noise (35.5%; $\beta = -4.77$, SE = 0.10, z = -45.51, $p < .001$), as well as a significant effect of AV status, with higher accuracy with Audio-Visual stimuli (75.1%) than Audio-Only stimuli (58.5%; $\beta = 0.72$, SE = 0.09, z = 7.93, $p < .001$). There was no main effect of or interactions with SOAE inhibition (all $p > .322$).

- Table 1 -

- Figure 1 -

- Figure 2 -

## Discussion

Our finding that there is increased attenuation for AV stimuli suggests that auditory input is less critical in the presence of visual input. Importantly, it suggests that top-down modulation of cochlear gain is not only elicited by the presence of noise and irrelevant, potentially distracting stimuli, but also by complimentary, yet redundant stimuli. The greater inhibition of SOAEs for the AV conditions also suggests that the observed effects on SOAEs were not merely a result of noise activation of the auditory efferent system (at the brainstem) but mediated by cortical processes dedicated to coordination across sensory modalities.

Consistent with past research (Zhao & Dhar, 2011), we found a reduction in SOAE amplitude in response to noise relative to quiet. One potential confound for this finding is the middle-ear muscle (MEM) reflex, which can have similar effects on SOAEs as efferent activity. While the MEM reflex can be actively monitored during the measurement of efferent inhibition of cochlear gain (see Deeter et al., 2009), our choice of SOAEs precluded this opportunity as SOAEs are generated without external stimulation. That said, the effects of the MEM reflex on SOAEs have been examined without external stimulation by using subjects who were able to voluntarily contract their MEM. The results, however, have been idiosyncratic and difficult to interpret (Burns et al.,1993). While some spontaneous emissions are attenuated by as many as 20 dB, others hardly demonstrate any change in level upon activation of the middle ear muscle reflex. Furthermore, the stimuli in the current study were presented at levels that are below typical MEM reflex thresholds (e.g., Mott et al., 1989; Zhao & Dhar, 2011), giving us confidence that the observed noise effects are likely due to MOC modulation. Lastly, while the MEM reflex could theoretically account for the noise effect, it could not explain the audio-visual effect, which was the primary focus of the current study. The auditory stimuli were identical in the audio-only and audio-visual conditions, and therefore would have elicited the same MEM response. And yet, we observed that there was greater inhibition of cochlear gain in the presence of complementary audio-visual inputs.

One potential limitation of the current experiment is the high proportion of women in the study sample (72%). Past research has found that the prevalence of SOAEs is significantly greater for women than men (Bilger, Matthies, Hammel, & Demorest, 1990), and so the generalizability of these effects to predominantly male populations will need to be addressed in future investigations. That said, our results were similar to the effect for transient evoked otoacoustic emissions (TEOAEs) demonstrated by Puel, Bonfils, and Pujol (1988) who reported a reduction in TEOAE, but not SOAE, amplitude while attending to a visual task. TEOAEs are elicited by presenting a click and recording the cochlear response, as opposed to SOAEs, which do not require external stimulation. To our knowledge, our study is the first to report an effect of attention on SOAEs (see Meric and Collet, 1994). The difference between our study and previous studies may be related to the type of tasks involved. In both Puel et al. and Meric and Collet, the auditory stimulus was irrelevant to the task and could be easily ignored. In our speech comprehension task, the auditory information was chosen to be meaningful and highly relevant, yet there was still a reduction in auditory gain in the presence of visual input. We propose that the redundancy between the auditory and visual information triggered the reduction in peripheral auditory gain. In conclusion, the results of this study suggest that auditory efferent responses can attenuate both distracting, irrelevant stimuli as well as complimentary, but redundant information. This pattern indicates that one function of top-down efferent control may be to optimize efficiency in processing multisensory stimuli.

## Acknowledgements

The authors thank Dr. Ken Grant for sharing his audio-visual speech stimuli and Dr. Jungmee Lee for her input on the initial design of this study. The authors would also like to thank Peter Kwak and Jaeryoung Lee for their assistance in recruiting participants and collecting data for this experiment. This project was funded in part by a training grant from the National Institute of Deafness and Other Communication Disorders T32-DC009399-04 to Tuan Lam and by grant RO1HD059858 to Viorica Marian.

## Author Contributions

V. M., S. D., and T. L. designed the study. T. L. collected the data. S. H. and T. L. analyzed the data and drafted the manuscript. V. M., S. D., and S.H. edited and finalized the manuscript. All authors contributed to interpretation of the results.

## Competing interests

The authors declare no competing financial interests.

# References

**Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

**Bilger, R. C., Matthies, M. L., Hammel, D. R., & Demorest, M. E.** (1990). Genetic implications of gender differences in the prevalence of spontaneous otoacoustic emissions. *Journal of Speech, Language, and Hearing Research, 33*, 418-432.

**Burns, E. M., Harrison, W. A., Bulen, J. C., & Keefe, D. H.** (1993). Voluntary contraction of middle ear muscles: effects on input impedance, energy reflectance and spontaneous otoacoustic emissions. *Hearing Research, 67*, 117-127.

**Davis, C., & Kim, J.** (1999). Perception of clearly presented foreign language sounds: the effects of visible speech. *In AVSP'99-International Conference on Auditory-Visual Speech Processing.*

**Deeter, R., Abel, R., Calandruccio, L., & Dhar, S.** (2009). Contralateral acoustic stimulation alters the magnitude and phase of distortion product otoacoustic emissions. *The Journal of the Acoustical Society of America*, *126*, 2413-2424.

**Grant, K. W., & Seitz, P. F.** (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*, 1197-1208.

**Grant, K. W., & Walden, B. E.** (1996). Evaluating the articulation index for auditory–visual consonant recognition. *The Journal of the Acoustical Society of America*, *100*, 2415-2424.

**Grant, K. W., Walden, B. E., & Seitz, P. F.** (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America*, *103*, 2677-2690.

**Guinan Jr., J. J.** (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear and Hearing*, *27,* 589-607.

**Harrison, W. A., & Burns, E. M.** (1993). Effects of contralateral acoustic stimulation on spontaneous otoacoustic emissions. *The Journal of the Acoustical Society of America*, *94,* 2649-2658.

**Honrubia, F. M., & Elliott, J. H.** (1968). Efferent innervation of the retina: I. Morphologic study of the human retina. *Archives of Ophthalmology*, *80,* 98-03.

**Hox, J.** (1998). Multilevel modeling: When and why. In Classification, data analysis, and data highways (pp. 147-154). Springer, Berlin, Heidelberg.

**Liberman, M. C., Liberman, L. D., & Maison, S. F.** (2014). Efferent feedback slows cochlear aging. *The Journal of Neuroscience*, *34*, 4599-4607.

**Jedrzejczak, W. W., Milner, R., Olszewski, L., & Skarzynski, H.** (2017). Heightened visual attention does not affect inner ear function as measured by otoacoustic emissions. *PeerJ, 5*, e4199.

**Judd, C. M., Westfall, J., & Kenny, D. A.** (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68,* 601-625.

**Kawase, T., Delgutte, B., & Liberman, M. C.** (1993). Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones. *Journal of Neurophysiology, 70*, 2533-2549.

**Kemp, D. T.** (1978). Stimulated acoustic emissions from within the human auditory system. *The Journal of the Acoustical Society of America, 64*, 1386-1391.

**Kumar, U. A., & Vanaja, C. S.** (2004). Functioning of olivocochlear bundle and speech perception in noise. *Ear and Hearing*, *25*, 142-146.

**Liberman, M. C., Liberman, L. D., & Maison, S. F.** (2014). Efferent feedback slows cochlear aging. *Journal of Neuroscience*, *34*, 4599-4607.

**Maison, S. F., & Liberman, M. C.** (2000). Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength. *The Journal of Neuroscience*, *20*, 4701-4707.

**McGurk, H., & MacDonald, J.** (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

**Meric, C., & Collet, L.** (1994). Attention and otoacoustic emissions: a review. *Neuroscience & Biobehavioral Reviews*, *18*, 215-222.

**Mott, J. B., Norton, S. J., Neely, S. T., & Warr, W. B.** (1989). Changes in spontaneous otoacoustic emissions produced by acoustic stimulation of the contralateral ear. *Hearing Rresearch, 38,* 229-242.

**Puel, J. L., Bonfils, P., & Pujol, R.** (1988). Selective attention modifies the active micromechanical properties of the cochlea. *Brain Research, 447*, 380-383.

**Sumby, W. H., & Pollack, I.** (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*, 212-215.

**Wagner, W., Frey, K., Heppelmann, G., Plontke, S. K., Zenner, H. P.** (2008). Speech-in-noise intelligibility does not correlate with efferent olivocochlear reflex in humans with normal hearing, *Acta Otolaryngologica*. *128*, 53-60.

**Walsh, K. P., Pasanen, E. G., & McFadden, D.** (2015). Changes in otoacoustic emissions during selective auditory and visual attention. *The Journal of the Acoustical Society of America*, *137*, 2737-2757.

**Zhao, W., & Dhar, S.** (2011). Fast and slow effects of medial olivocochlear efferent activity in humans. *PLoS One*, *6*, e18725.

**Zhao, W., & Dhar, S.** (2012). Frequency tuning of the contralateral medial olivocochlear reflex in humans. *Journal of Neurophysiology*, *108*, 25-30.

**Table 1**

*Average change in SOAE peak level for audio-only and audio-visual conditions*

| | **Audio-Only** | **Audio-Visual** |
|---|---|---|
| **Noise** | -4.56 dB | -4.71 dB |
| **Quiet** | -2.09 dB | -2.28 dB |

## Figure Legends

*Figure 1.* Spectrum of an ear canal recording from a participant exhibiting multiple spontaneous otoacoustic emissions (SOAEs). Level and frequencies of multiple SOAEs recorded from one subject. The large (and stable) spontaneous emissions that were monitored for this study are marked by vertical arrows.

*Figure 2.* Average reduction in peak level in response to audio-only and audio-visual speech. Change in SOAE level (dB; during stimulus presentation minus baseline) in response to audio-only speech and the additional reduction for audio-visual speech averaged across all SOAE frequencies. The bar on the left represents speech in quiet while the bar on the right represents speech in noise. Error bars represent the standard error of the mean for each condition.