Identifying Hearing Deficiencies from Statistically Learned Speech Features for Personalized Tuning of Cochlear Implants

Bonny Banerjee[1,2], Lisa L. Mendel[3], Jayanta K. Dutta[1,2], Hasti Shabani[1,2], Shamima Najnin[1,2]

[1]Institute for Intelligent Systems, [2]Dept. of Electrical and Computer Engineering, [3]School of Communication Sciences and Disorders

The University of Memphis, Memphis, TN 38152, USA

bbnerjee@memphis.edu

Abstract

Cochlear implants (CIs) are an effective intervention for individuals with severe-to-profound sensorineural hearing loss. Currently, no tuning procedure exists that can fully exploit the technology. We propose online unsupervised algorithms to learn features from the speech of a severely-to-profoundly hearing-impaired patient round-the-clock and compare the features to those learned from the normal hearing population using a set of neurophysiological metrics. Experimental results are presented. The information from comparison can be exploited to modify the signal processing in a patient's CI to enhance his audibility of speech.

The Problem

Hearing loss is the most common birth defect in the U.S. and amounts to billions in societal losses over a lifetime. Cochlear implants (CIs) are FDA-approved hearing devices that are surgically implanted into the inner ear. CIs are an effective intervention for adults and children with severe-to-profound sensorineural hearing loss, helping them gain the ability to hear, achieve age-appropriate reading skills, and develop communication skills comparable to their hearing counterparts, in a safe and reliable way. Personalized tuning of the parameters of a CI's signal processing strategy based on hearing deficiencies is vital for successful outcomes. Without proper tuning, optimal access to sound cannot be delivered, even in the case of good candidate selection, surgery, and rehabilitation support.

In the current state-of-the-art, personalized tuning of a CI to optimize the hearing sensations received is a challenging and time-consuming task, even for highly trained and experienced audiologists, due to four reasons.

1. *Large number of tunable parameters*. The signal processing in a CI is controlled by roughly 200 tunable parameters with nonlinear interdependencies. The number of tests to evaluate an individual's hearing increases combinatorially with the number of parameters and their domain size, leading to the curse of dimensionality.
2. *Paucity of data*. Each test requires on average 90 seconds, which contributes one point to an individual's unknown high-dimensional error surface in the parametric space. An audiologist can test a patient for only a few points which is inadequate to estimate the optimal one.
3. *Substantial noise and variability in each patient's data*. Pure tone and speech audiometry in conjunction with subjective feedback from patients are the primary outcome measures used to assess the effectiveness of CI tuning. Unfortunately, these measures assess a fraction of one's auditory performance and provide little analytical feedback to the fitter. Consequently, many audiologists rely on instantaneous feedback from patients' subjective judgment which is variable and inconsistent and may not reflect the best device settings for their speech recognition (Gordon et al., 2004). As patients are often very young or may never have heard 'normally' before, this feedback relates more to comfort than to the intrinsic accuracy of sound coding (Govaerts et al., 2010).
4. *Wide variation in hearing loss characteristics among patients*. The characteristics of hearing loss vary widely among individuals based on different factors, such as age, cause of hearing loss, degree of hearing loss, pre- or post-lingual deafness, and so on. Hence, fine-tuning procedures ought to be personalized in order to be effective.

Currently a number of semi-automated tuning procedures exist, such as optimizing the correlation between objective measures for optimal MAP prediction and their behavioral equivalents (McKay et al., 2005), heuristics in the form of deterministic rules applied in an ad hoc manner to all patients (Govaerts et al., 2010), and ensemble methods for judiciously selecting the points for personalized testing (Banerjee and Krause, 2013). These approaches show considerable variability and fail to accurately capture the hearing characteristics of an individual. Two factors

may be attributed for this limitation—lack of adequate data and the analysis of a patient's stimulus-response errors in terms of handcrafted features, such as Jakobson et al.'s (1961) distinctive features.

Solution Approach

Our working hypothesis is that the deficiencies in hearing for people with severe-to-profound hearing loss are reflected in their speech (Ryalls et al., 2003). It is assumed that our algorithms can reside in the CI device and tune it internally, as proposed by Krause et al. (2010), thereby having continuous access to the user's speech. We overcome the paucity of data by learning features from the speech output data around the clock. We address the issue of handcrafted features by learning the features for each user in an unsupervised and online manner.

The proposed approach to personalized tuning of a CI consists of three steps: learn features from the speech of the CI user round-the-clock in online and unsupervised manner, compare these features to those learned from the speech of normal hearing population using a set of neurophysiological metrics to identify hearing deficiencies, and exploit this information to modify the signal processing in the user's CI to enhance his audibility of speech. These steps are to be executed adaptively, allowing enough time for the user to adapt to the new parameter setting.

Feature learning. We propose a two-layered neural network architecture for investigating and comparing multiple encoding and feature learning strategies. We consider the winner-take-all and sparse coding strategies for encoding, and stochastic gradient descent and block coordinate descent for learning. Combinations of these along with different network topologies lead to different objectives, such as clustering, clustering by ignoring outliers, sparse coding and predictive coding, resulting in different kinds of features. Further, learning from different representations of audio is investigated, such as time-amplitude series, spectrogram and cochleogram.

Identifying hearing deficiencies. Our two-layered network is conceptualized as the hair cell layer in the inner ear of the brain's auditory pathway. Each feature learned by a higher layer neuron may be conceived as representing the audio pattern encoded in the receptive field of a unique hair cell in the cochlea. Evaluation of statistically learned audio features is a challenge as there does not exist well-defined metrics for comparing their properties to those of receptive fields of auditory cells. In order to facilitate comparison, we algorithmically define five metrics—distribution of characteristic frequencies, equal loudness contour, tuning curve, skewness and Q_{10} value of a tuning curve—for audio features. These metrics have been used extensively in neurophysiology (while directly recording from individual or groups of cells) but were unavailable to the feature learning community. Our experimental results demonstrate the utility of the five metrics by comparing, with a physiological model (Zilany et al., 2014), the following five sets of audio features—puretones, gammatone filters, features learned using clustering, clustering by ignoring outliers, and sparse coding—from three classes of audio data, namely, speech (male and female), music and natural sounds. If the computation by hair cells can be assigned an objective, this comparison allows us to discover that objective function.

Our experiments with subjects reveal the differences in the metrics between a subject with severe-to-profound hearing loss and the normal hearing population. Deficiencies in hearing are manifested in these differences. For example, lack of characteristic frequencies in a particular frequency range indicates a dead region in the cochlea and wide tuning curves in a frequency range are indicative of poor discrimination in that range. Exploiting this information to modify the signal processing in the user's CI to enhance his audibility of speech is left as future research.

Acknowledgment

The research reported in this paper was supported by NSF CISE Grant No. 1231620.

References

Banerjee, B., and Krause, L. S. 2013. Automatic performance optimization for perceptual devices. U.S. patent no. 8,401,199.

Gordon, K., Papsin, B. C., and Harrison, R. V. 2004. Programming cochlear implant stimulation levels in infants and children with a combination of objective measures. *International Journal of Audiology*, 43 Suppl 1:S28-32.

Govaerts, P. J., Vaerenberg, B., De Ceulaer, G., Daemers, K., and Schwauwers, K. 2010. Development of a software tool using deterministic logic for the optimization of cochlear implant processor programming. *Otology & Neurotology*, 31(6):908–918.

Jakobson, R., Fant, C. G. M., and Halle, M. 1961. *Preliminaries to speech analysis: The distinctive features and their correlates*. MIT Press, Cambridge, MA.

Krause, L. S., Shrivastav, R., Banerjee, B., and Holmes, A. E. 2010. System and methods for remotely tuning hearing devices. U.S. patent application no. 12/748,819.

McKay, C. M., Fewster, L., and Dawson, P. 2005. A different approach to using neural response telemetry for automated cochlear implant processor programming. *Ear Hear*, 26(4):38S-44S.

Ryalls, J., Larouche, A., and Giroux, F. 2003. Acoustic comparison of CV syllables in French-speaking children with normal hearing, moderate-to-severe and profound hearing impairment. *Journal of Multilingual Communication Disorders*, 1(2):99-114.

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. *Journal of the Acoustical Society of America*, 135(1):283-286.