PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Opening the Black Box: Real-Time Speech Perturbation Experiments Reloaded

Bahne Hendrik BAHNERS*[1]; Sebastian HEIDELBERG[2]; Joseph BAADER[2];
Ruben VAN DE VIJVER[3]; Markus BUTZ[1]; Julian ROHRHUBER[2]

[1] Institute of Clinical Neuroscience and Medical Psychology, Heinrich-Heine University, Düsseldorf, Germany

[2] Institute for Music and Media, Robert Schumann Hochschule, Düsseldorf, Germany

[3]Institute of Linguistics and Information Science, Heinrich-Heine University, Düsseldorf, Germany

ABSTRACT

In neuroscience, speech perturbation experiments are a well-studied and useful mean to assess speech pathologies and neuronal integration mechanisms. These experiments often depend on complex signal processing technology. In order to enable replication and modification of experiments as well as the interpretation of their results, technological details need to be accessible. However, in practice the experiment's actual mechanism often remains hidden in black boxes like digital signal processors or other audio equipment. We conducted a study with 20 Parkinson patients, assessing vocal responses to pitch shifted AF. While we could replicate earlier experimental findings of Parkinsonian speech pathologies, we introduced an open-source and easy-applicable setup for real-time speech perturbation experiments. It runs on standard audio interfaces, allows researchers to interactively reprogram the signal flow at runtime and can be applied both inside and outside the laboratory.

Keywords: speech perturbation, open-source, neuroscience

1. INTRODUCTION

Speech is an essential aspect of everyday life. Consequently, any dysfunction affecting speech quality has major social implications for a patient. Therefore, a better understanding of the complex mechanisms underlying these dysfunctions is of high interest to the subjects of speech pathology and neuroscience (1,2).

As a complex human behavior, speech requires the fast coordination of various subsystems (3). In order to perform this complex task, an elaborate control system is necessary, which continuously integrates sensory and motor information (4). The requirements for this control system of speech are easily met by healthy individuals. Unsurprisingly, many neurodegenerative diseases like Parkinson's Disease (PD) and Alzheimer's Disease (AD) show an influence on the sensorimotor integration of speech (5,6). Especially, the change of auditory feedback (AF) perception seems to play a major role for the altered function of auditory-motor integration (AMI) mechanisms (1).

To study the role of AMI for speech pathologies, speech-perturbation experiments can be utilized, employing online manipulation of AF (6). By introducing frequency changes into AF, neurophysiological as well as vocal responses are provoked (5). Whenever the motor cortex sends a speech command and an efference copy, the auditory and sensorimotor systems feed back information about the speech performance to the brain. The actual performance is then compared to the motor prediction coded by the efference copy and vocal output is adjusted accordingly (7). AD and PD patients notably compensate much stronger to altered AF than healthy individuals in these kinds of experiments (1).

Another typical observation among PD patients lies in the wider pitch variability in vowel

* bahne.bahners@gmail.com

vocalizations. Interestingly, this pitch variability turned out to be positively correlated with the stronger compensations (i.e. pitch response magnitudes) to AF, mentioned above (1). Several methods for online speech-perturbation have been applied in this field (1,2,8). However, the actual underlying acoustic mechanisms of signal manipulation are difficult to report, as signal processing is hidden inside digital signal processors (DSP) or other audio equipment.

Also, reproductivity is limited due to the fact, that most control scripts are custom-made and inaccessible. The need for expensive proprietary technology, like specialized DSP is an obstacle to experimental practice.

We therefore aimed at designing a novel method, which is transparent, flexible and reproducible:

(a) *transparent*, in the way that its process runs on any standard audio interface,

(b) *flexible*, by using live-coding methods that allow the change of both parameters and program even during the experiment,

(c) *reproducible*, as it is open-source, available online for free and can be installed easily on any experimental computer.

2. METHODS

2.1 Atomicity, Literacy, Liveness

For being able to deeply reconfigure the audio processing system at runtime, a flexible and concise programming system is required. The SuperCollider programming language (9,10) together with its built-in extensions for live coding is a suitable environment for this purpose. It is open source, well maintained by a large user community and has been applied in scientific contexts, in particular for sonification (11). Its basic architecture combines a high-level interpreted programming language with a reliable, low latency signal processing server, which is controlled via the open sound control (OSC) protocol. Our design followed the principles of liveness, atomicity, and literacy:

a) *liveness*: the development should happen within a system that is always running. For reproducibility, the final setup is frozen for the experiment itself.

b) *atomicity*: runtime changes to the system remain maximally independent from the rest of the system.

c) *literacy*: core technology is given in publishable and readable format (12).

The SuperCollider programming language, with the live coding capabilities of its subsystem JITLib (13), provides the necessary infrastructure for liveness, atomicity and literacy. Live coding is a relatively new programming paradigm, in which programming is an integral part of the runtime of a program (14). Rather than already being completed at the time of execution, the primary interface with the running process is code. This reduces the need for graphical user interfaces, allows for independently and interactively rewriting parts of the program, and simplifies the program text for readability (15). Instead of separating development and application steps, the system may thereby be left open to the experimenter.

2.2 The Pitch Shifting Paradigm

The experiment requires a change in pitch of a voice signal in real-time without simultaneously changing the intraspectral relations. Not changing these relations reduces the number of potential error sources and maintains the patient's recognition of her or his own voice. However, all known pitch shifting paradigms manipulate these spectral relations in so far as they maintain a real-time signal. To resolve this contradiction, pitch shifting accounts for different kinds of trade-offs. Solutions in the frequency domain, like the Phase Vocoder (16), struggle with reconstructing the correct temporal relations of partial frequencies in the process of inverse Fourier transform. Solutions in the time domain, like Synchronous Overlap and Add (17), have problems with phase continuity when reattaching temporal bins. More modern approaches are often blackboxed or depend on powerful hardware which can handle the necessary computations in real-time, and they are not designed with scientific applications in mind. This interferes with our goal to create a transparent and comprehensible system that runs on simple setups.

The vocalization of a single vowel produces a homogeneous and repetitive signal on a constant base frequency. Each pitch shifting episode in our experiment lasted only 200 milliseconds. These particular circumstances allowed us to use a technique similar to a tape delay. The signal was recorded live into a short buffer. During the pitch shifting episode, its play-back rate was reduced to an

appropriate lower rate using cubic interpolation, replacing the live signal by the pitch shifted version. At the end of the pitch shifting episode the modified signal was cross-faded back to the live signal over a period of 100 milliseconds.

This simple method is easily replicable in SuperCollider using our provided source code (https://github.com/musikinformatik/pspeech), but also reconstructable in other programming languages. The open and modular structure of our implementation additionally allows for the replacement, fine tuning, and direct comparison of pitch shifting methods at run-time. The system can thus fluently adapt to unexpected requirements and new experimental settings.

2.3 Controlling the pitch shifting activation

Experiments may be tedious for the patients and experimental setups are inherently error prone. This means that care should be taken to make the procedure perceptually smooth, reliable, sufficiently short, and not overly repetitive. Moreover, to avoid post hoc data cleaning as far as possible, the experimental setup should be designed to only collect data when operated correctly. To achieve this, we performed a real time analysis of the voice signal for stability of voicing measuring amplitude, spectral flatness and autocorrelation (Fig.1), so that the pitch shift is only triggered under the right conditions.

Adequate cross-fading of all parameters avoids noise bursts both in regular work and in development. The pitch shifting itself is triggered within a bounded random interval, as to avoid habituation and strain. The activation signal is recorded to a sound file just like the voice signals, as to keep their relation fully traceable.

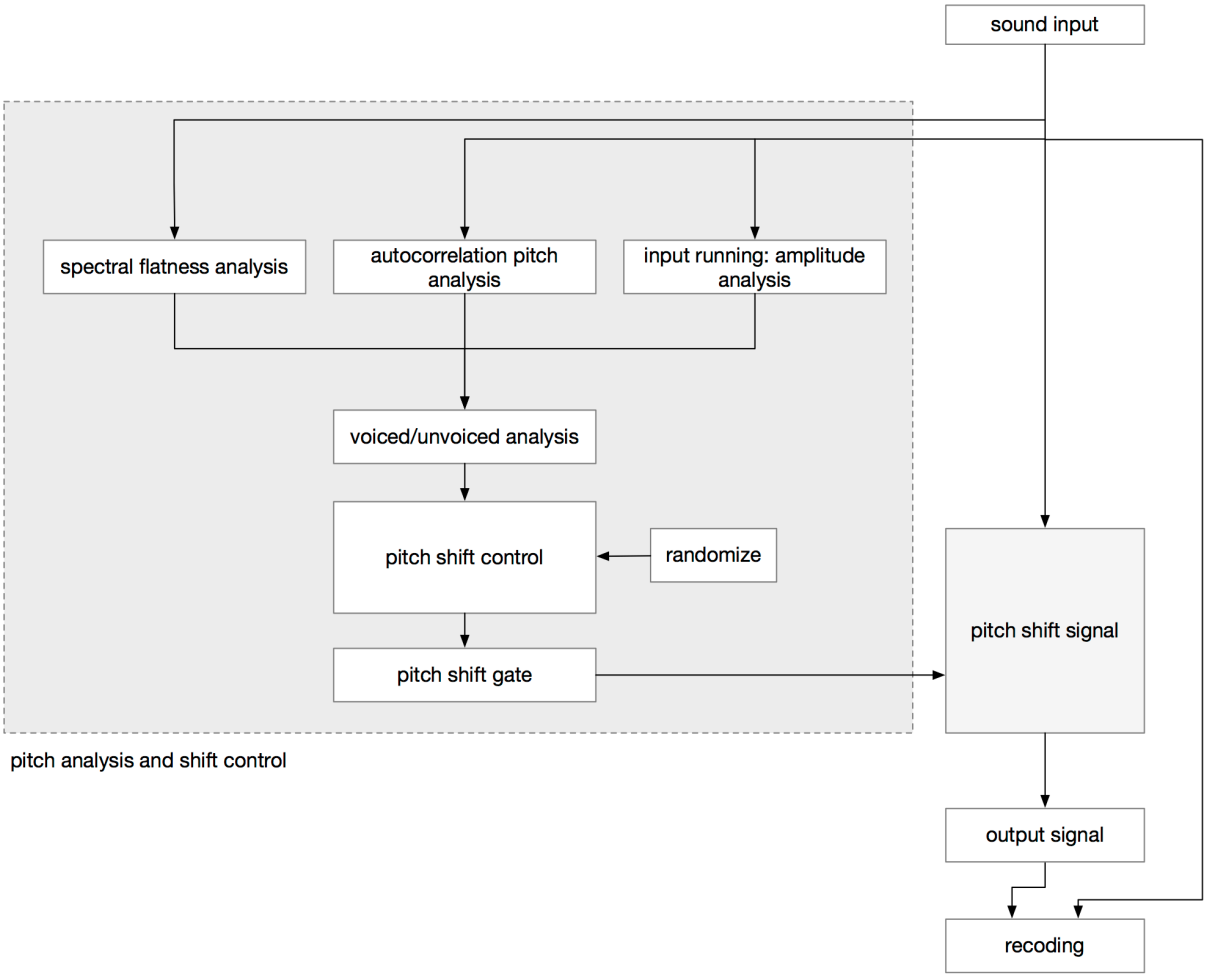

Figure 1 – real time voice analysis and pitch shifting procedure

2.4 Experiment

Twenty German speaking PD patients (15 male, 5 female; 62.4 ± 6.7 years) were recruited during their annual control visit at the Centre for Movement Disorders and Neuromodulation at the University Hospital Düsseldorf. The mean disease duration after appearance of the first movement symptoms was 9.6 ± 4.37 years. The patients were implanted with a deep brain stimulation (DBS) system in the subthalamic nucleus (STN) 24.9 ± 21.3 months ago and had a significant therapeutic effect regarding motor scores of the unified Parkinson's Disease rating scale (UPDRS) (off stimulation: 27.85± 11.9 vs. on stimulation: 15.15±6.46, $t=6.018$, $df=19$, $p<0.001$).

An optical microphone (Sennheiser MO 2000) was installed at a distance of 5 centimeters to the patient's mouth. After the signal was processed on the experimental computer's built-in audio interface (SoundMX integrated Digital HD, Intel Corporation©, 64 bits; 33 MHz), AF was played back to the participant through insert earphones (ER-1, Etymotic Research Inc.) via a mixing console (Behringer© XENYX 502 PA) with an overall maximum delay of 10 milliseconds. The built-in audio interface had a hardware delay of 5.4 milliseconds itself and recordings were sampled at 96kHz. A visual presentation with experimental instructions was executed on another computer independent from sound-processing (MacBook Pro 2,5 GHz Intel Core i5) and shown to the patients over a rear projection system.

While his or her stimulation had been turned off for at least 30 minutes, the patient was asked to vocalize the German vowel "E" [e]. A visible count-down from 3 to 1, lasting approx. 3 seconds, was shown on the screen to prepare the patient for the vocalization. Afterwards, a blue circle containing the letter "E" appeared, indicating the beginning of the first vocalization period. The circle disappeared clockwise within 5.6 seconds. A white screen signified a pause that lasted 5 to 9 seconds, including the count-down for the next vocalization period. These pauses between each of the 30 vocalization periods in total became longer towards the end of the experiment with the goal to prevent vocal fatigue.

During each vocalization period the patient's voice was pitch shifted downwards 200 cents for 200 milliseconds 3 to 4 times. In reaction to the artificial pitch change, the patients compensated with a pitch change themselves (Fig.2). Pitch shifting happened randomly 500 to 1000 milliseconds after speech onset with random inter-stimulus intervals of 700 to 900 milliseconds. We recorded the altered as well as the unaltered signals. After each block of 30 vocalizations we played back the altered

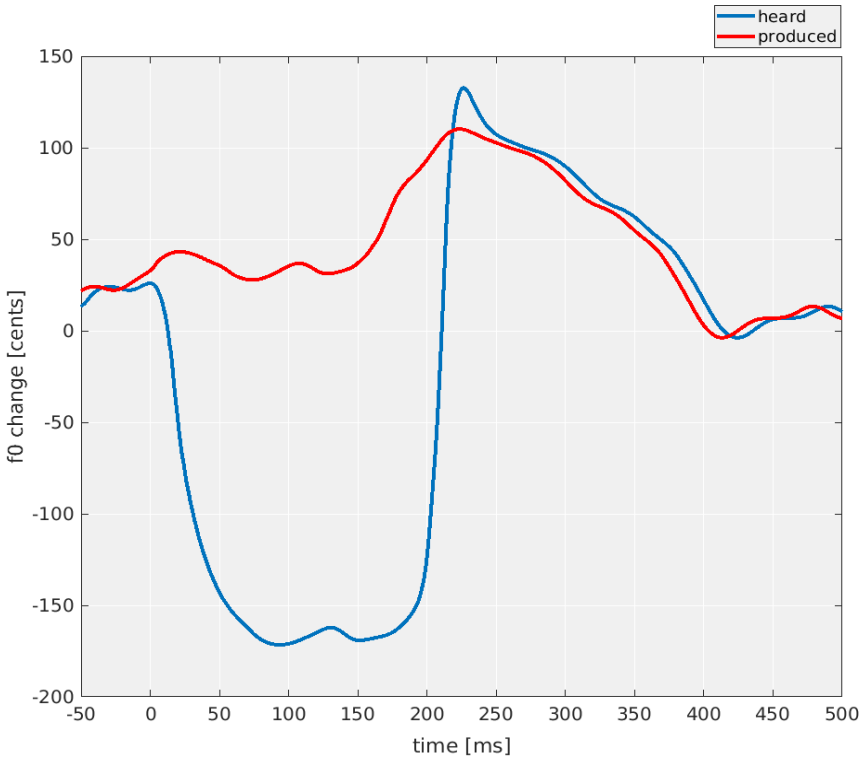

Figure 2 – heard f0 changes and produced compensating f0 responses

recording including the pitch shifted sequences to the patient. The time of each pitch shifting onset was saved in a separate audio file with amplitude changes from 0 to 1. These time points were used to extract the time locked pitch response contours of each trial.

2.5 Analysis

To analyze vocal compensating responses to AF we extracted the individual pitch contours of every subject's recording in PRAAT (18). The pitch contours were transferred to the cent scale.

We then extracted the responses time locked to the pitch shifting onset, using f0 values 100 milliseconds before and 500 milliseconds after the onset. By rejecting trials with negative response magnitude values, we assured that only opposing, i.e. compensating responses were considered for the response analysis. We calculated the response magnitude by subtracting the mean baseline f0 (100 milliseconds before pitch shifting onset) from the maximum f0 value in a time window of 100 to 300 milliseconds after the pitch shifting. We made this calculation for each trial and averaged the response magnitudes.

The extraction of responses as well as trial sorting and the other calculations including statistics were conducted with MATLAB (2018b, MathWorks Inc.). We used a Spearman's rank-order correlation test, due to the fact that the vocal response magnitudes were non-normally distributed (Shapiro-Wilk normality test: $p<0.05$).

3. RESULTS

To proof the quality of sound manipulation and validate our experimental setup, we had to show that we were able to replicate earlier findings and to provoke actual vocal responses among subjects, known to elicit especially strong compensating responses, i.e. Parkinson patients.

Indeed, PD patients in our study showed relatively strong responses to f0 perturbations. The mean f0-response magnitude was 23.91±10.44 cents (standard deviation). These response magnitudes are, in fact, similar to the results of equivalent experiments, where e.g. Huang et al. observed a mean response magnitude of 24±13 cents in patients with PD, while healthy individuals showed response magnitudes of 15±5 cents to a pitch shifting of 200 cents (1). Other studies with the same parameters could find response magnitudes for healthy subjects in a similar range (19).

Earlier findings looked at a specific interaction between pitch responses and pitch variability (1). We could find a similarly strong correlation (Fig. 3), which was highly significant ($r=0.726$, $p<0.001$).

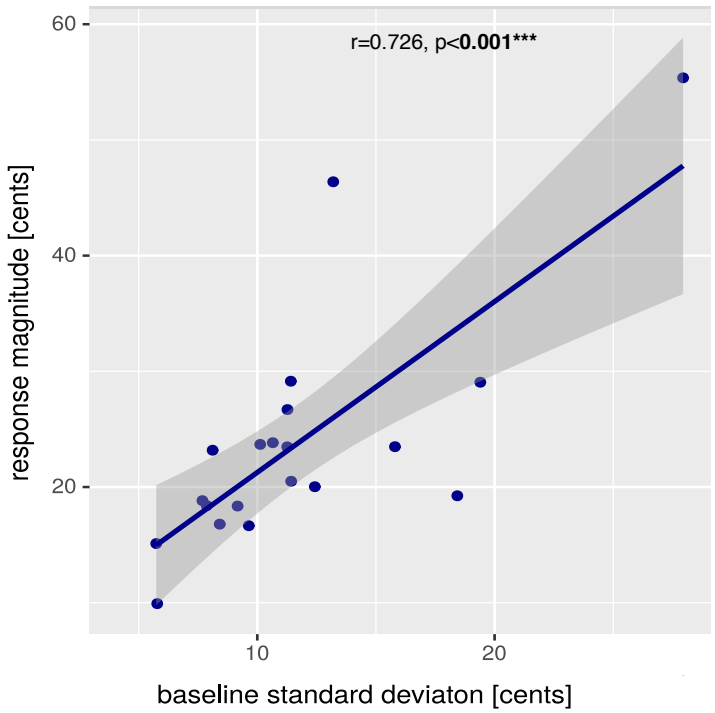

Figure 3 – f0 response magnitudes correlate with f0 baseline standard deviation

4. DISCUSSION

While some components of every experiment remain black boxed for good reasons, it is important to give full access at least to those components which are active part of the current research. Today, much signal processing can be done on standard hardware and with high level open source programming languages. Instead of relying on opaque devices, our novel experimental setup documents the precise conditions under which it was applied. Moreover, due to its modular character, it does not only solve one single problem, but serves as a prototype for a whole class of experiments. Using this prototype could enable scientists to start projects with low technical and financial barriers.

Concerning scientific research questions, the setup can easily be used for different types of auditory stimulus presentation with low latencies and jitter. All stimulus properties are documented and can be adjusted dynamically, which makes it easy to replicate scientific studies. Due to its timing accuracy, the setup can serve as a means to study dynamic processes of speech production, rather than only simple and repetitive speech stimuli.

Also, the setup can be installed on any computer using a standard audio interface. Therefore, it can also be brought outside the laboratory. Possible applications may lie in point of care testing (POCT) in neurological hospitals, where patients could be examined in a bed-side assessment. Audio files of speech recordings can be used offline for standard speech diagnostics in PRAAT (18). Using speech perturbation in diagnostics might even help to detect speech pathologies with higher sensitivity.

5. CONCLUSION

The replication of scientific studies is important to validate previous research. By introducing an open-source and easily applicable solution for real-time speech perturbation, we hope to increase transparency, flexibility and reproducibility within neuroscientific signal processing.

Our study showed why and in how far experimental practice may benefit from an inclusion of programming into the experimental protocol. Its flexible nature opens up new application possibilities concerning speech diagnostics.

ACKNOWLEDGEMENTS

Special thanks to Holger Krause who helped integrating the system into the laboratory setup and to Marika Schulz for her support with UPDRS ratings. We thank all patients, who participated in this study.

REFERENCES

1. Huang X, Chen X, Yan N, Jones JA, Wang EQ, Chen L, et al. The impact of parkinson's disease on the cortical mechanisms that support auditory–motor integration for voice control. Hum Brain Mapp. 2016 Dec 1;37(12):4248–61.
2. Behroozmand R, Karvelis L, Liu H, Larson CR. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clin Neurophysiol Off J Int Fed Clin Neurophysiol. 2009 Jul;120(7):1303–12.
3. Hixon TJ, Weismer G, Hoit JD. Preclinical Speech Science: Anatomy, Physiology, Acoustics, and Perception, Third Edition. Plural Publishing; 2018. 759 p.
4. Franken MK, Eisner F, Acheson DJ, McQueen JM, Hagoort P, Schoffelen J-M. Self-monitoring in the cerebral cortex: Neural responses to small pitch shifts in auditory feedback during speech production. NeuroImage. 2018 Oct 1;179:326–36.
5. Ranasinghe KG, Gill JS, Kothare H, Beagle AJ, Mizuiri D, Honma SM, et al. Abnormal vocal behavior predicts executive and memory deficits in Alzheimer's disease. Neurobiol Aging. 2017 Apr 1;52:71–80.
6. Mollaei F, Shiller DM, Baum SR, Gracco VL. Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. Brain Res. 2016 Sep 1;1646:269–77.
7. Kort NS, Nagarajan SS, Houde JF. A bilateral cortical network responds to pitch perturbations in speech feedback. NeuroImage. 2014 Feb 1;86:525–35.
8. Tourville JA, Cai S, Guenther F. Exploring auditory-motor interactions in normal and disordered speech. Proc Meet Acoust. 2013 May 14;19(1):060180.
9. McCartney J. Rethinking the computer music language: SuperCollider. Comput Music J. 2002;(26):61–68.

10. Wilson S, Cottle D, Collins N. SuperCollider Book. Wilson S, Cottle D, Collins N, editors. MIT Press; 2008.
11. Bovermann T, Rohrhuber J, de Campo A. Laboratory Methods for Experimental Sonification. In.
12. Knuth DE. Literate Programming. Stanford, California: Center for the Study of Language and Information; 1992. (CSLI Lecture Notes, no. 27.).
13. Rohrhuber J, de Campo A, Wieser R. Algorithms today - Notes on Language Design for Just In Time Programming. In: Proceedings of International Computer Music Conference. Barcelona: ICMC; 2005. p. 455–8.
14. Blackwell A, McLean A, Noble J, Otto J, Rohrhuber J. Collaboration and learning through live coding (Dagstuhl Seminar 13382). Dagstuhl Rep. 2014;3(9):130–168.
15. Rohrhuber J, de Campo A. Just In Time Programming. In: Collins N, Wilson S, Cottle D, editors. The SuperCollider Book. Cambridge, Massachusetts: MIT Press; 2011.
16. Flanagan JL, Golden RM. Phase Vocoder. Bell Syst Tech J. 1966;45(9):1493–1509.
17. Zue VW. Digital Processing of Speech Signals, by L. R. Rabiner and R. W. Schafer. J Acoust Soc Am. 1980 Apr 1;67(4):1406–7.
18. Boersma P. Praat, a system for doing phonetics by computer. Glot Int [Internet]. 2002 [cited 2018 Feb 16];5. Available from: http://dare.uva.nl/search?arno.record.id=109185
19. Li J, Hu H, Chen N, Jones JA, Wu D, Liu P, et al. Aging and Sex Influence Cortical Auditory-Motor Integration for Speech Control. Front Neurosci. 2018;12:749.