# PHYSIOLOGICAL BASES OF THE ENCODING OF SPEECH EVOKED FREQUENCY FOLLOWING RESPONSES

[1]Nike Gnanateja G., [2]Ranjeet Ranjan, & [3]Sandeep M.

**Abstract**

*Many studies have shown that fundamental frequency (F0) is represented in the speech evoked Frequency Following Response (FFR), however, it is not clear as to what aspect of the stimulus is the basis for the F0 coding. The energy at the Fo alone is less likely to be the basis, as our ear is less sensitive to very low frequencies which is evident from the very high RETSPL. Thus, the present study was taken up to analyse the independent role of high frequency harmonics and stimulus envelope in the encoding of the speech evoked FFRs. In the Experiment 1 of the present study, FFRs were elicited with a high-pass filtered syllable and compared with that of the unfiltered syllable. Results showed that the FFRs elicited for the 2 stimuli were not different in spectrum. This finding implies that the FFRs are primarily coded by the higher harmonics (frequencies beyond F2), and the lower harmonics contribute less for the coding of F0 and the first formants. However, as the envelope was same in both the stimuli, it cast a doubt that the responses obtained were because of the envelope and not the lower harmonics. To verify this, Experiment 2 was carried out, wherein FFRs were recorded to stimuli with envelope of the vowel portion removed without altering the fine structure. The FFRs elicited by the fine-structure stimulus revealed that the F0 amplitude was significantly lower compared to the original stimulus which implies that envelope is the key parameter for the coding of FFRs.*

**Key Words:** *Brainstem pitch encoding, envelope following responses, fine-structure encoding*

Auditory brainstem responses (ABR) elicited with speech stimuli have received much attention in the last two decades. The ABR to a speech stimulus has basically three main components, the onset response, the frequency following response (FFR) and the offset response (Greenberg, 1980; Galbraith Arbagey, Branski, Comerci, & Rector, 1995; Skoe & Kraus, 2010). The FFR has received considerable attention in the last decade. The FFR is a sustained auditory evoked potential which mimics the oscillations in the stimulus and has been found to be originating primarily from the inferior colliculi and the rostral brainstem structures (Cunningham, Nicol, Zecker, Bradlow, & Kraus, 2008; Greenberg, Marsh, Brown, & Smith, 1987; Krishnan & Gandour, 2009; Russo, Nicol, Musacchia, & Kraus, 2004). The FFRs have been recorded for stimuli like, vowels, consonant-vowel syllables, low frequency tones, modulated tones and also to instrumental music. The consonants, sparing the continuants, in speech do not contribute much for the FFR, rather they elicit onset responses. On the other hand, sustained vowels elicit FFRs that mimic the stimulus.

There has been considerable research on the types of stimuli that are most suitable for recording the FFRs, the results of which have thrown light on the physiological bases of the FFRs. Dau (2003) demonstrated that FFRs to low frequency tones, represent synchronized brainstem activity mainly stemming from mid and high-frequency excitation of the basilar membrane, and not from units tuned to frequencies around the signal frequency. They hypothesized that the temporal envelope conveyed by the higher frequency regions might be more responsible for the generation of the FFR than the characteristic frequency itself. Contrary to Dau (2003), Greenberg, Marsh, Brown, & Smith (1987) recorded FFRs for missing fundamental stimuli with varied stimulus envelopes and demonstrated that the FFRs are not a result of a phase locking to the stimulus envelope. Thus, discrepancies exist in literature about the bases for coding of FFRs.

**Need and specific aim of the study**

The syllable /da/ of 40 millisecond duration having a fundamental frequency of nearly 105 Hz and five formants has been extensively used to record the speech evoked FFR (Krizman, Skoe, & Kraus, 2010; Abrams, Nicol, Zecker, & Kraus, 2006; Hornickel, Skoe, & Kraus, 2009). Most of these studies have quantified the spectrum of FFR in terms of its amplitude at fundamental frequency (F0) and the first two formant (F1 and F2) frequency ranges. Although, it is clear from these studies that Fo is represented in the elicited FFR, it is not clear as

[1]Research Officer, All India Institute of Speech and Hearing (AIISH), Mysore-06, Email: nikegnanateja@gmail.com, [2]Research Officer, AIISH, Mysore-06, Email: ranjanbs3@yahoo.co.in & [3]Lecturer in Audiology, AIISH, Mysore-06, Email: msandeepa@gmail.com

to what aspect of the stimulus is the bases for the F0 coding. The energy at the Fo is less likely to be the bases as our ear is less sensitive to very low frequencies, which is evident from the very high Reference Equivalent Sound Pressure Level. Psychophysical theories of pitch perception have demonstrated that it is the higher harmonics which determine the pitch of the signal (Plomp, 1967; Moore & Peters, 1992). Objective correlates of this have been demonstrated by Greenberg et al. (1987) and Chambers, Feth and Burns (1986) using harmonic complexes where they showed that the harmonics in the signal help represent the fundamental in the FFR. They further concluded that the FFR follows the fine structure alone and is not related to the waveform envelope. However a close look at the waveforms in their study shows considerable difference in the amplitude of the FFR with changes in the stimulus envelope. Hence, the role of stimulus envelope, along with the fine structure was suspected to be the bases for FFR and this needed experimental investigation.

Thus, the present study was taken up to analyse the independent role of higher harmonics and stimulus envelope on the speech evoked FFRs.

## Method

The study was conducted as two experiments. Experiment 1 was conducted to analyse the role of the higher harmonics in the speech evoked FFRs while the Experiment 2 was conducted to analyse the role of the stimulus envelope in the speech evoked FFRs and to help explain the results of Experiment 1.

### Experiment 1

#### Participants

Thirteen adults in the age range of 18 to 24 years participated in the first experiment. All the participants had normal hearing sensitivity, middle ear functioning and speech perception in noise, on preliminary evaluations. Preliminary evaluations included puretone audiometry, immittance evaluation and, the assessment of speech identification scores at 0 dB signal-to-noise ratio. The participants had pure tone hearing thresholds of 15 dBHL or lesser, in the octave frequencies between 250 to 8 kHz on air conduction testing. They had Type-A tympanogram with normal acoustic reflex thresholds ruling out the presence of any middle ear pathology (Jerger, Anthony, Jerger, & Mauldin, 1974). Their speech identification scores were more than 60% in both ears, in the presence of speech noise.

#### Test Stimuli

The main stimulus used in the study was a synthetic 40 msec /da/ syllable, same as the one used by Abrams, Nicol, Zecker, and Kraus (2006). A filtered derivative of the /da/ stimulus was prepared by high pass filtering the original /da/ stimulus using a sixth order butterworth filter with a cut-off frequency of 1700 Hz in Adobe Audition, version 3.0. The cut-off frequency corresponded to the second formant frequency of the signal. The waveforms and spectra of the two stimuli are shown in Figure 1. The frequency following responses for the original /da/ stimulus and the filtered /da/ stimulus were elicited with a stimulation rate of 10.9/s at an intensity level of 80 dBSPL presented in alternating polarity.

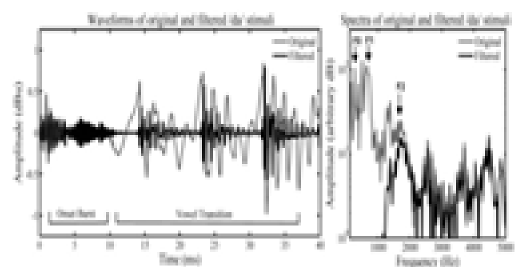

Figure 1*: The waveforms and the spectra of the original and the filtered /da/ stimuli*

#### Test Procedure

The actual test procedure involved recording of the speech evoked ABR using different stimuli. The participants were comfortably seated on a reclining chair. Before starting the recording, low absolute and relative electrode impedance were ensured. Biologic Navigator Pro (version 7) was used to record the FFRs. The stimulus locked responses were acquired between -11.2 to +52.8 msec in a vertical (Cz-C7) electrode montage. The responses elicited by 2000 alternate polarity stimuli were averaged and replicated. The stimuli were presented at 10.9/s rate through ER-3A insert phones at 80 dBSPL.

The waveforms were spectrally analysed using Brainstem toolbox 2010 (Skoe & Kraus, 2010). The amplitudes at the fundamental frequency (F0 = 103-121 Hz), first formant frequency (F1 = 454-719 Hz) and second formant frequency (F2 = 721-1155 Hz) were analysed in the region of 11 to 45 milliseconds. The FFR spectral amplitudes thus obtained, were scaled into arbitrary decibel values and compared across the different stimuli.

## Results

All the statistics in the data were performed using Statistical Package for Social Sciences, version-17 (SPSS, V.17). The data was analysed on the Kolmogorov Smirnov which showed that the data was normally distributed. Multiple paired t-tests were used to compare the spectral amplitudes at F0, F1, and F2 of the FFRs elicited by original /da/ and filtered /da/. The mean and standard deviation of the spectral amplitudes for the two stimuli are given in Table 1. Results of paired t-test (Table 2) revealed no significant differences between the amplitudes at F0, F1 and F2 for the FFRs elicited by the original and filtered stimuli.

Table 1: *Mean and standard deviation of F0, F1, and F2 amplitudes (arbitrary dB) for the FFRs elicited by original /da/ and filtered /da/ and the results of paired t-test*

| Parameter | Stimulus /da/ | Mean (N = 13) | SD | t | df | p |
|---|---|---|---|---|---|---|
| F0 | Original | 5.69 | 2.25 | 0.76 | 12 | 0.47 |
| | Filtered | 5.23 | 2.62 | | | |
| F1 | Original | 0.70 | 0.34 | 1.17 | 12 | 0.27 |
| | Filtered | 0.85 | 0.22 | | | |
| F2 | Original | 0.29 | 0.12 | 0.78 | 12 | 0.45 |
| | Filtered | 0.32 | 0.07 | | | |

## Experiment 2

### Participants

Fifteen participants in the age range of 18 to 27 years participated in this experiment. The subject selection criteria were the same as that in Experiment 1.

### Test Stimuli

The Hilbert envelope of the vowel portion of the original /da/ stimulus was extracted. The stimulus was then divided by its Hilbert envelope so as to obtain the fine structure of the vowel. This was achieved using a customized script on a MATLAB platform (version 7.14). Thus, the new stimulus was same as the original stimulus in terms of the burst portion till the vowel onset, and dissimilar in terms of the envelope of the vowel portion. This new stimulus was operationally termed fine-structure /da/. The waveforms and the spectra of the original and the fine-structure /da/ are shown in Figure 2.

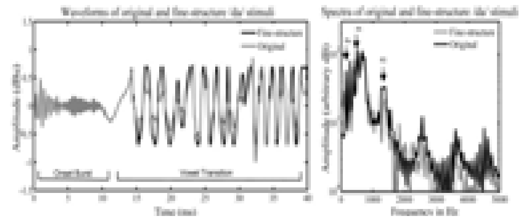

Figure 2*: Waveforms and spectra of original /da/ and fine-structure /da/.*

The stimulus and recording parameters, as well as analysis methods were same as that used in Experiment 1. Kolmogorov-Smirnov test revealed that the data in all the parameters were normally distributed. The spectral amplitudes of the FFRs at F0, F1 and F2 were compared between the stimuli on multiple paired t-tests.

## Results

The mean and standard deviation of the spectral amplitudes of FFRs elicited by the two stimuli are given in Table 3. Table 3 also gives the results of paired t-test. Results revealed large statistically significant difference ($p<0.000$) between the amplitudes at F0 for the FFRs elicited by the original and fine-structure /da/. There was also a smaller, however, significant difference in the ampitudes at F1 ($p = 0.024$) and F2 ($p = 0.033$) between the two stimuli.

Table 3: *Mean and standard deviation of amplitudes (arbitrary dB) at F0, F1, and F2 for the FFRs elicited by original /da/ and fine-structure/da/, and the results of paired t-test*

| Parameter | Stimulus /da/ | Mean (N = 13) | SD | t-value | df | Level of significance |
|---|---|---|---|---|---|---|
| F0 | Original | 6.94 | 2.86 | 5.38 | 14 | 0.00 |
| | Fine-structure | 2.72 | 1.55 | | | |
| F1 | Original | 1.69 | 0.63 | 2.53 | 14 | 0.02 |
| | Fine-structure | 1.47 | 0.49 | | | |
| F2 | Original | 0.47 | 0.47 | 2.36 | 14 | 0.03 |

## Discussion

The onset responses were not analysed as it was not the focus of the study. The FFRs elicited by the filtered stimulus were similar to those elicited by the original stimulus as can be seen from Figure 3. Also there was no difference in the FFRs based on the spectral analysis. Thus, the elimination of the lower harmonics in the /da/ stimulus did not affect the FFRs appreciably.

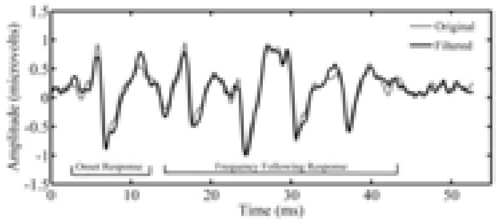


Figure 3: *Grand average FFRs elicited by the original /da/ and filtered /da/.*

This representation of F0 and lower formants in the FFRs, while the low frequency spectral components were removed, may be because of two probable mechanisms: (1) as proposed by the temporal models of pitch perception (Terhardt, Stoll, & Seewan, 1974), the decoding of the harmonic relationship in the higher harmonics at the brainstem level may be aiding to code F0, F1 and F2. (2) The residual low frequency envelope information from the higher harmonics helped code the FFRv as shown in Figures 4 and 5.

Experiment 2 was carried out to verify which of these two explanations holds good. The envelope of the vowel portion of the stimulus was smeared to obtain the fine structure of the vowel portion, while the burst portion of the stimulus was left unaltered.

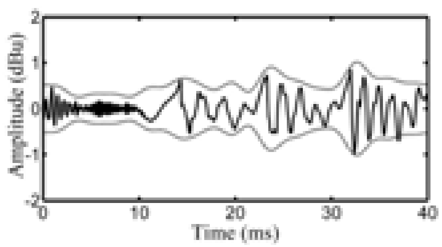


Figure 4*: Waveform of the original /da/ (black) with schematic representation of the stimulus envelope (red).*

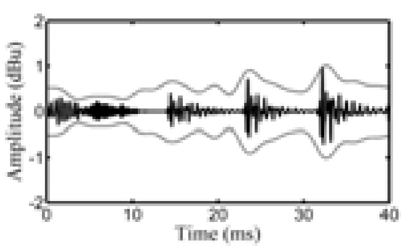


Figure 5*: Waveform of the filtered /da/ (black) with schematic representation of the stimulus envelope (red) of original /da/*

The FFRs elicited by the fine-structure /da/ showed that the Fo amplitude was significantly lower compared to those elicited by the original /da/. The burst portion in the fine-structure /da/ stimulus elicited onset responses exactly similar to the onset responses elicited by the original /da/. However, the FFRs elicited by the fine-structure /da/ were considerably different in the morphology and the number of peaks observed. The grand averaged FFRs for the two stimuli can be seen in Figure 6.

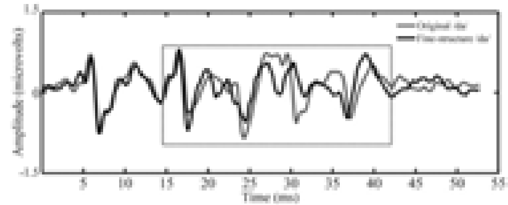


Figure 6: *Grand average FFRs elicited by the original /da/ and the fine structure /da/ stimuli*.

The grey marked area shows the FFR region which varied for the two stimuli. The results suggest that the strikingly similar FFRs to the two stimuli in Experiment 1 can be attributed to the similarity in the envelope between the two stimuli used as can be seen from Figure 4 and Figure 5. This finding gives strong evidence to the second explanation proposed, i.e. the FFRs are basically coded by the stimulus envelope as also proposed by Gardi, Merzenich, and McKean (1979) and Stillman, Crow, and Moushegian (1987). Additionally, the first experiment suggests that the lower harmonics do not help significantly in the coding of the FFRs. The lower harmonics are 'resolved' and the combined output of the resolved harmonics from the cochlea has been demonstrated to have poor temporal coherence (Dau, 2003) and thus did not contribute to the coding of the FFRs which are primarily the envelope following responses. However, the higher harmonics being 'unresolved' produce a temporally coherent cochlear output and help in coding of theses envelope following responses.

## Conclusions

The F0, F1 and F2 information represented in the FFRs are primarily the result of phase locking to the stimulus envelope mediated by the precise temporal coding of higher harmonics. However, the role of the lower harmonics and the energy at the fundamental frequency itself appears to be somewhat limited. From these findings, one can also infer that energy below F2 need not be synthesized while generating the stimulus for a study that aims to track the brainstem encoding of pitch.

## Acknowledgements

The authors would like to thank the Director, All India Institute of Speech and Hearing for providing the necessary infrastructure for the study.

## References

Abrams, A. A., Nicol, T., Zecker, S. G., & Kraus, N.(2006). Auditory Brainstem timing predicts cerebral asymmetry for speech. *The Journal of* Neuroscience, 26(43), 11131-11137.

Chambers, R., Feth, L., & Burns, E. (1986). The relationship between the human frequency following response and the low pitch of complex tones. *Journal of the Acoustical Society of America, 75*, S81(A).

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning impairment: deficits and strategies for improvement. *Clinical Neurophysiology, 112*,758-767.

Dau, T. (2003). The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *Journal of the Acoustical Society of America, 113*, 936-50.

Galbraith, G. C., Arbagey, P. W., Branski, R., Comerci, N., & Rector, P. M. (1995). Intelligible speech encoded in the human brainstem frequency following response. *Neuroreport, 6*, 2363-2367.

Gardi, J., Mezenich, M., & McKean, C. (1979). Origins of the scalp-recorded frequency-following response in the cat. *Audiology*, *18*, 358-381.

Greenberg, S. (1980). Temporal neural coding of pitch and vowel quality. *Working Papers in phonetics. WPP-52*. Department of Linguistics, UCLA

Greenberg, S., Marsh, J.T., Brown, W.S. & Smith, J.C. 1987. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing Research, 25*, 91-114.

Hornickel, J., Skoe, E. & Kraus, N. (2009). Sub-cortical laterality of speech encoding. *Audiology and Neuro-otology*, *14*, 198-207.

Jerger, J., Anthony, L., Jerger, S., & Mauldin, L. (1974). Studies in impedance audiometry, III: Middle ear disorders. *Archives of Otolaryngology, 99*, 409-413.

Jewett, D. L., Romano, M. N., & Williston, J. S. (1970). Human auditory evoked potentials: possible brainstem components detected on the scalp. *Science*, 167, 1517-1518

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain & Language*, *110*, 135-148

Krizman, J., Skoe, E., & Kraus, N. (2010). Stimulus rate and subcortical auditory processing of speech. *Audiology & Neurotology*, *15*, 332-342.

Moore, B. C. J., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *Journal of the Acoustical Society of America, 91*, 2881-2893.

Plomp, R. (1967). Pitch of complex tones. *Journal of Acoustical Society of America, 41,* 1526-1533.

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 115*, 2021-2030.

Skoe, E., & Kraus, N. (2010). Auditory brainstem responses to complex sounds: A tutorial. *Ear & Hearing, 31*(3), 302-324.

Stillman, R. D., Crow, G., & Moushegian, G. (1978). Components of the frequency-following potential in man. *Electroencephalography and Clinical Neurophysiology*, *44*, 438-446.

Terhardt, E., Stoll, G., & Seewann, M. (1982). Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. *Journal of the Acoustical Society of America, 71,* 671–678.