

# DENTOFACIAL DISHARMONY PATIENTS' SIBILANTS DIFFER FROM CONTROLS' MORE IN SOURCE THAN FILTER PROPERTIES

Madeleine Oakley<sup>1</sup>, Auvi Tran<sup>2</sup>, Ciana Paye<sup>1</sup>, Emma Trudan<sup>1</sup>, Timothy Turvey<sup>2</sup>, George Blakey<sup>2</sup>,  
David Zajac<sup>2</sup>, Jeff Mielke<sup>1</sup>, Laura Jacox<sup>2</sup>

<sup>1</sup>North Carolina State University; <sup>2</sup>University of North Carolina Chapel Hill  
mjoakley@ncsu.edu; ljacox@live.unc.edu

## ABSTRACT

This study uses multitaper spectral analysis to examine the differences in consonants produced by patients who present with different dentofacial disharmonies (DFD) including severe overbites (Class II), underbites (Class III) and anterior open bites. Previous studies have found that patients with these malocclusion types all produce sibilants and plosives with increased spectral center of gravity and increased spectral spread relative to controls. This result is puzzling since some DFD groups differ from controls in opposite ways. To better understand the articulatory basis of these differences, we apply several spectral shape measures and find that all groups of DFD patients produce /s ʃ t tʃ/ with mid-frequency spectral peaks that are less prominent than those of the control group, but peak frequency measures are largely the same across all groups. This indicates that the DFD patients differ more in sibilant noise source than front cavity size.<sup>1</sup>

**Keywords:** sibilant, multitaper, disordered speech, malocclusion, source-filter

## 1. INTRODUCTION

The great majority of patients who have dentofacial disharmonies (DFD) present with speech sound disorders [1, 2, 3]. DFD patients present with skeletal and dental discrepancies in the anterior-posterior (AP, horizontal), vertical, and/or transverse dimensions, including Class II overbite, Class III underbite, and anterior open bite (AOB). Previous work examining the relationship between DFDs and speech sound disorders has found an association between malocclusion and speech disorders [4, 1].

Studies examining the acoustic properties of disordered speech have shown that there are differences in spectral properties of consonants between DFD patient groups. Using spectral moment analysis, significant increases were seen in the 1st and 2nd spectral moments (center of gravity

and spectral spread) of DFD patients including Class II, III and AOB groups when compared to Class I reference subjects for /k t ʃ s tʃ/ [5, 6]. Patients with AOB presented with the greatest quantitative differences and Class II with the smallest. Among the Class III DFD cohort, the subgroup of Class III AOB patients experienced the greatest differences in the first and second spectral moments compared to the control group, potentially due to the combination of vertical and AP discrepancies [5, 6].

These spectral moment results are surprising as each DFD cohort has unique jaw disproportions and vocal tract anatomy, and it is not predicted that all DFD groups would demonstrate increased first and second spectral moments. One reason for the discrepancy between predicted acoustic differences between groups and articulations may be the use of spectral moment analysis. As discussed by [7], changes in certain moment measurements can be affected by multiple articulatory movements, which may be obscuring differences across DFD cohorts.

Spectral moment analysis typically involves applying the discrete Fourier transform to a waveform. Because the resulting spectra have a large variance, these estimates may have poor precision when applied to fricative waveforms [8]. They are suitable for measuring spectral moments but less so for measuring properties of spectral peaks which may offer more insight into the articulatory basis of the observed differences. Multitaper spectral analysis averages multiple estimators based on the Discrete Fourier Transform and detection of harmonic components despite presence of external noise; this helps to produce a single spectrum minimizing spectral bias, noise, variance, and error [9, 10]. Such spectra are more suitable for more targeted measurements of spectral peaks. Measures such as mid-frequency spectral peak frequency, change in the mid-frequency peak over time, the amplitude difference between the low frequency minimum and the mid-frequency peak, and high frequency slope measures [7]. These acoustic

measurements more closely reflect articulatory changes, such as labiality (which lowers mid-frequency peaks), changes in the size of the front cavity over time, and aerodynamic effects of the decrease in the oral constriction over time.

The current study applies multitaper spectral analysis to speech recordings from pre-surgical DFD populations to evaluate how the severity of Class II, Class III, and AOB malocclusions correlate with speech distortion of consonants. This data provides insight into the physiologic interplay between articulators and explores how jaw disproportions are linked to speech disorders.

## 2. METHODS

### 2.1. Participants

One hundred fifty DFD patients with Class II closed bite ( $n = 34$ ), Class III closed bite ( $n = 64$ ), Class I anterior open bite ( $n = 7$ ), Class II anterior open bite ( $n = 9$ ), or Class III anterior open bite ( $n = 36$ ) malocclusions were consecutively enrolled from a DFD clinic. All patients were screened by an orthodontist prior to enrollment based on their malocclusions. Fifty-three reference controls with Class I occlusion and skeletal base were included. All patients were aged 18-40 years old, 115 were female, and 88 were male. Orthodontic and surgical records included occlusal measurements, dental models, photos (intraoral and extraoral), and panoramic and cephalogram radiographs.

### 2.2. Materials and Procedure

The target sounds /s j t tj k/ were selected for analysis based on the prediction that malocclusions affect the production of these consonants. 20 English words containing word-initial /s j t tj k/ before the vowels /i u æ a/ were embedded in the carrier phrase “say \_\_ again.” All phrases were randomized and presented one at a time. The words were each repeated 3 times for a total of 12 target tokens per phone per speaker. Recordings were made in a sound-attenuated booth while participants wore a head-mounted microphone.

### 2.3. Analysis

The Montreal Forced Aligner [11] was used to segment and align the recordings using transcripts. Stop and affricate releases were identified based on the high frequency energy and voicing probability [12], roughly following [13]. Spectra were measured from 20 ms windows aligned to the middle

of each /s j/ frication interval and to the start of the release of each /t tj k/ using the spectRum package [14] for R [15], with 8 tapers and a bandwidth parameter of 4. Preemphasis was not applied.

This study focuses on two acoustic measurements for each consonant which are designed to access different articulatory parameters, following [16, 7, 17]. The frequency of the mid-frequency spectral peak (measured in the 2-5kHz range for /j tj/ and in the 3-7kHz range for /s t/) is expected to be associated with the size of the front cavity between the lingual constriction and the mouth opening. Spectral amplitude difference (the difference in amplitude between the mid-frequency spectral peak and a lower-frequency spectral trough (measured in the 1-2kHz range for /j tj/ and otherwise in the 1-3kHz range) is expected to be associated with the noise source itself. /k/ lacks a mid-frequency spectral peak so it was not measured in this way.

Separate mixed-effects linear regression models were run for each consonant with the different acoustic measurements as the dependent variable. Patient group and gender are included as fixed-effects, and word and speaker are included as random effects.

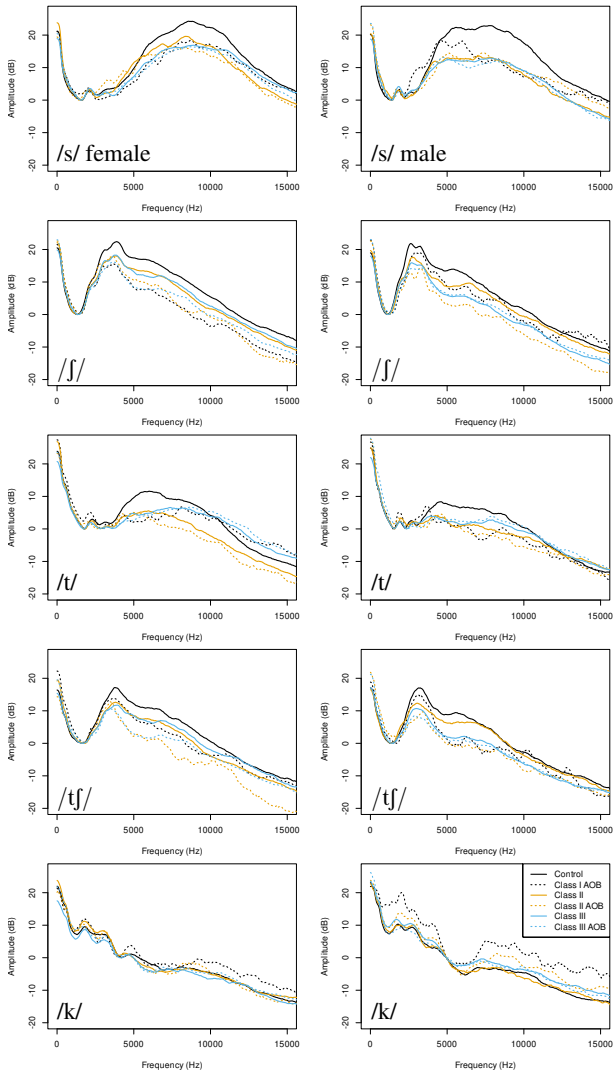
## 3. RESULTS

The mean spectra for each consonant for each patient group and gender, are shown in Figure 1. The spectra are adjusted so that 0dB corresponds to the low-frequency trough. For /s j t tj/, the control group has more prominent spectral peaks, and the frequencies of these peaks appear similar across groups. The /k/ spectra appear similar across patient groups, with the exception of male Class I AOB, which is probably a spurious difference due to the small number of patients in that category ( $n = 2$ ).

### 3.1. Spectral amplitude difference

Spectral amplitude difference is expected to be correlated with the noise source. As seen in Figure 1, there appears to be a tendency for the control group to produce all of the consonants (except /k/) with a higher amplitude than the other patient groups. This measure was compared across patient groups for each target consonant. Overall, the spectral amplitude difference for /s j t tj/ is higher for the control group than for the other patient groups. There is also a tendency for the AOB groups to have a lower spectral amplitude difference compared to non-AOB groups. These findings are represented in Figure 2, divided by gender groups.

Separate regression models for each consonant

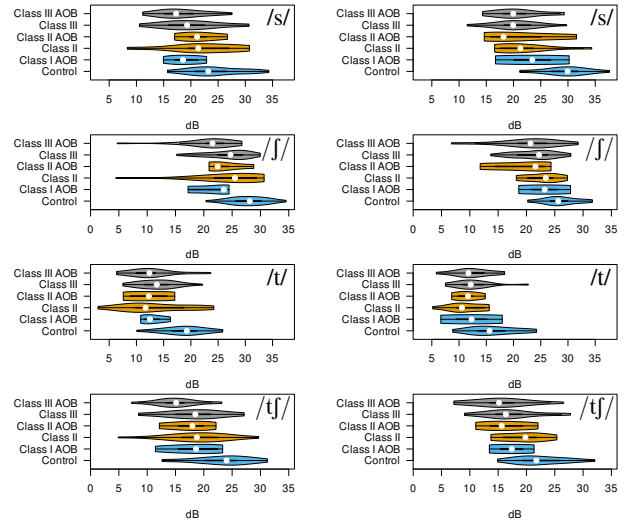


**Figure 1:** Mean spectra averaged for each patient group for each consonant, by gender (left: female; right: male)

confirm that each patient group has a significant effect on spectral amplitude difference for /s ʃ t tʃ/. The outputs of the regression models are summarized in Table 1, showing the effect of each patient group on the spectral amplitude difference for each consonant.

### 3.2. Peak frequency

The frequency of the mid-frequency peak is expected to correlate with front cavity size. This measure was compared across patient groups for all four consonants with mid-frequency peaks, as shown in Figure 3. The control group does not differ in peak frequency value from the different patient groups for the post-alveolar consonants /ʃ/ or /tʃ/. The control group does, however, produce



**Figure 2:** Spectral amplitude difference (dB) measured for each patient group for each consonant with a mid-frequency peak, by gender (left: female; right: male)

Group	/s/	/ʃ/
Control	25.63 $p = 4.48e-16^{***}$	27.73 $p < 2e-16^{***}$
Class I AOB	-6.44 $p = 0.002^{**}$	-6.15 $p = 0.0008^{***}$
Class II	-4.35 $p = 8.37e-05^{***}$	-3.28 $p = 0.0004^{***}$
Class II AOB	-5.48 $p = 0.001^{**}$	-4.83 $p = 0.0007^{***}$
Class III	-6.54 $p = 4.10e-11^{***}$	-4.11 $p = 5.41e-7^{***}$
Class III AOB	-7.45 $p = 1.08e-10^{***}$	-6.38 $p = 8.17e-11^{***}$

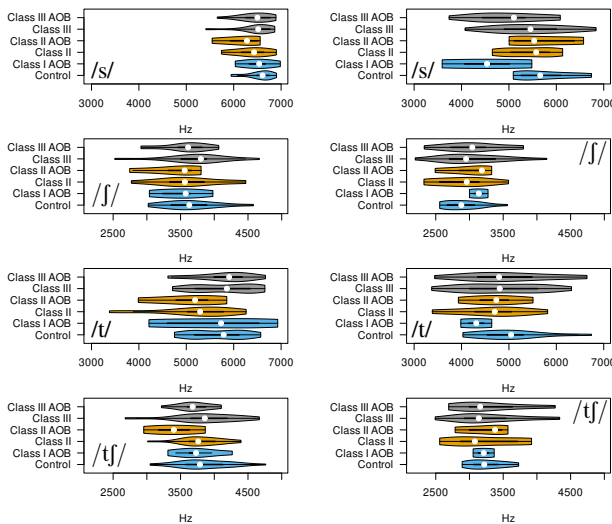
Group	/t/	/tʃ/
Control	18.40 $p = 2.92e-7^{***}$	23.15 $p < 2e-16^{***}$
Class I AOB	-5.77 $p = 0.001^{**}$	-7.44 $p = 0.0002^{***}$
Class II	-5.15 $p = 1.79e-8^{***}$	-3.83 $p = 0.0001^{***}$
Class II AOB	-6.04 $p = 1.12e-5^{***}$	-6.28 $p = 6.77e-5^{***}$
Class III	-4.28 $p = 5.42e-8^{***}$	-5.14 $p = 1.48e-8^{***}$
Class III AOB	-5.13 $p = 2.57e-8^{***}$	-7.78 $p = 8.60e-13^{***}$

**Table 1:** Estimates from linear regression models for each consonant, showing the effect of patient group on spectral amplitude difference (dB)

/s/ with a slightly higher frequency spectral peak (measured at the full frequency range) than Class I AOB and Class II patient groups. The control group also produces /t/ with a higher peak frequency than all other patient groups (measured in the lower-mid frequency range). Separate linear regressions for each consonant confirm these observations, showing a significant effect of patient group type for /s t/, but not /ʃ tʃ/. The output of the regressions for /s t/ are summarized in Table 2.

## 4. DISCUSSION

This study provides detailed quantitative assessments on a large DFD sample representing all major malocclusion classifications. The multitaper analysis shows differences across DFD groups that



**Figure 3:** Mid-frequency peak (Hz) measured for each patient group for each consonant with a mid-frequency peak, by gender (left: female; right: male)

Group	/s/	/t/
Control	8623 $p = 3.66e-12^{***}$	4265 $p = 5.93e-14^{***}$
Class I AOB	-993.5 $p = 0.048^*$	-957.9 $p = 0.0003^{***}$
Class II	-600.4 $p = 0.018^*$	-645.6 $p = 2.82e-6^{***}$
Class II AOB	-235.9 $p = 0.54$	-546.8 $p = 0.008^{**}$
Class III	-60.16 $p = 0.78$	-514.7 $p = 1.40e-5^{**}$
Class III AOB	-415.0 $p = 0.1$	-600.4 $p = 1.44e-5^{***}$

**Table 2:** Estimates from linear regression models, showing the effect of patient group on peak frequency (Hz) for each consonant with a significant difference

have not previously been observed.

First, the results of this study show that the control group produces /s ʃ t tʃ/ with a higher spectral amplitude difference than any other patient group. As discussed by [7] this measurement is designed to measure the ‘sibilance’ of fricatives and reflect variations in the noise source at different frequency regions. With the control group having the largest spectral amplitude followed by Class II non-AOB, and then all AOB, this could mean that Class II non-AOB patients are best able to posture so as to produce an alveolar constriction resulting in turbulence most similar to that produced by Class I controls. Most Class II patients present with deficient mandibles in which their tongues are also naturally positioned more posteriorly, which contributes to a more similar constriction location compared to the controls. The AOB cohort has a more difficult time positioning the tongue tip and teeth to produce control-like turbulence. The Class III patients exhibiting a lower amplitude as well can be explained by those patients commonly

having prognathic mandible in which their tongues naturally rest more anteriorly. Adaptation for tongue position in Class III patients is difficult, potentially leading to a more anterior constriction location. The results of this study confirm that this measurement does reflect differences in articulatory configuration, as it is predicted that each DFD cohort has different jaw configurations that effect the constriction location of each target consonant.

Second, the control group does *not* produce post-alveolar consonants with a different peak frequency from other patient groups, but *does* produce /s/ with a higher peak frequency than Class I AOB and Class II groups, and *does* produce /t/ with a higher peak frequency than all other patient groups. Postalveolar consonants are produced with lip rounding in English, and it appears that all patient groups are able to achieve a degree of lip rounding that yields peak frequencies similar to controls when lip rounding is a feature of the consonant and/or when the target peak frequency is low. It appears that all patient groups are able to compensate with their lips, which is a movement available to anyone, for the rounded consonants. When lip rounding is not a feature of the consonant, as in /t/, all patient groups have somewhat lower peak frequencies than controls. Class I AOB and Class II have lower peak frequencies for /s/ as well, which is also unrounded. Note that the compensatory behavior that would raise peak frequency is *less* lip rounding. Patient groups do not seem to be using lip unrounding to achieve the higher peak frequency values for unround consonants.

Previous studies [5, 6] found higher center of gravity and spectral spread for various consonants produced by all DFD groups. The fact that center of gravity was observed to be *higher* is likely explained by the flatter spectra produced by Class II, Class III, and all AOB groups. Notably, those previous FFT-based analyses applied preemphasis and calculated spectral moments over a wide frequency range (0-17.64kHz), meaning that a relatively flat spectrum can yield a higher center of gravity than one with a prominent spectral peak under 10kHz. Using multitaper spectra to identify the mid-frequency spectral peak associated with the front cavity resonance made it possible to determine that the main difference between Controls and DFD cohorts is that the controls produce more prominent spectral peaks, indicative of greater sibilance produced at the lingual-alveolar constriction. Differences in peak frequency observed for /t s/ could be attributable to differences in the noise source that cannot be easily compensated for with the lips.

## 5. REFERENCES

- [1] L. I. Black, A. Vahratian, and H. J. Hoffman, "Communication disorders and use of intervention services among children aged 3-17 years: United states, 2012. nchs data brief. number 205." *Centers for Disease Control and Prevention*, 2015.
- [2] A. Ocampo-Parra, B. Escobar-Toro, V. Sierra-Alzate, Z. V. Rueda, and M. C. Lema, "Prevalence of dyslalias in 8 to 16 year-old students with anterior open bite in the municipality of envigado, colombia," *BMC oral health*, vol. 15, no. 1, pp. 1–6, 2015.
- [3] M. A. Morris, S. K. Meier, J. M. Griffin, M. E. Branda, and S. M. Phelan, "Prevalence and etiologies of adult communication disabilities in the united states: Results from the 2012 national health interview survey," *Disability and health journal*, vol. 9, no. 1, pp. 140–144, 2016.
- [4] R. Feeney, L. Desha, J. Ziviani, and J. M. Nicholson, "Health-related quality-of-life of children with speech and language difficulties: A review of the literature," *International Journal of Speech-Language Pathology*, vol. 14, no. 1, pp. 59–72, 2012.
- [5] M. M. B. Keyser, H. Lathrop, S. Jhingree, N. Giduz, C. Bocklage, S. Couldwell, S. Oliver, K. Moss, S. Frazier-Bowers, and C. Phillips, "Impacts of skeletal anterior open bite malocclusion on speech," *FACE*, vol. 3, no. 2, pp. 339–349, 2022.
- [6] H. Lathrop-Marshall, M. M. B. Keyser, S. Jhingree, N. Giduz, C. Bocklage, S. Couldwell, H. Edwards, T. Glesener, K. Moss, S. Frazier-Bowers *et al.*, "Orthognathic speech pathology: Impacts of class iii malocclusion on speech," *European Journal of Orthodontics*, vol. 44, no. 3, pp. 340–351, 2022.
- [7] L. L. Koenig, C. H. Shadle, J. L. Preston, and C. R. Mooshammer, "Toward improved spectral measures of/s: Results from adolescents," 2013.
- [8] P. F. Reidy, "A comparison of spectral estimation methods for the analysis of sibilant fricatives," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. EL248–EL254, 2015.
- [9] D. B. Percival and A. T. Walden, *Spectral analysis for physical applications*. Cambridge University Press, 1993.
- [10] C. H. Shadle *et al.*, "Acoustics and aerodynamics of fricatives," *The Oxford handbook of laboratory phonology*, pp. 511–526, 2012.
- [11] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Proc. Interspeech 2017*, 2017, pp. 498–502.
- [12] S. Gonzalez and M. Brookes, "Pefac-a pitch estimation algorithm robust to high levels of noise," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 518–530, 2014.
- [13] J. Cronenberg, M. Gubian, J. Harrington, and H. Ruch, "A dynamic model of the change from pre-to post-aspiration in andalusian spanish," *Journal of Phonetics*, vol. 83, 2020.
- [14] P. F. Reidy, "spectRum," 2013, r package. [Online]. Available: <https://github.com/patrickreidy/spectRum>
- [15] R Core Team, "R language definition," 2015.
- [16] L. Jesus and C. Shadle, "A parametric study of the spectral characteristics of European Portuguese fricatives," vol. 30, pp. 437–464, 2002.
- [17] C. H. Shadle, W.-r. Chen, L. L. Koenig, and J. L. Preston, "Acoustic variability of fricatives in normal adults," 2020, poster presented at ISSP 2020, Yale University.

---

<sup>1</sup> This research was funded by the National Institutes of Health (NIH), specifically the National Institute of Dental and Craniofacial Research (NIDCR), through a K08 award (to Laura Jacox), with a Grant Award Number 1K08DE030235- 01A1 and by the National Science Foundation through grant SMA-1730479 (to Jeff Mielke).