

## Research Article

# Multitaper Spectrum Analysis of Consonants Produced by Patients With Dentofacial Disharmonies

Auvi Tran,<sup>a</sup> Madeleine Oakley,<sup>b</sup> Ciana Payne,<sup>c</sup> Emma Trudan,<sup>d</sup> Nare Ghaltakhchyan,<sup>a</sup> Timothy Turvey,<sup>a</sup> George Blakey,<sup>a</sup> David Zajac,<sup>a</sup> Jeff Mielke,<sup>b</sup>  and Laura Anne Jacox<sup>a</sup> 

<sup>a</sup>Adams School of Dentistry, University of North Carolina at Chapel Hill <sup>b</sup>Department of English, North Carolina State University, Raleigh <sup>c</sup>Department of STEM Education, North Carolina State University, Raleigh <sup>d</sup>Department of Communication Sciences and Disorders, North Carolina Central University, Durham

## ARTICLE INFO

## Article History:

Received May 4, 2023

Revision received October 3, 2023

Accepted November 17, 2023

Editor-in-Chief: Cara E. Stepp

Editor: Raymond D. Kent

[https://doi.org/10.1044/2023\\_JSLHR-23-00280](https://doi.org/10.1044/2023_JSLHR-23-00280)

## ABSTRACT

**Purpose:** This study investigates differences in American English consonants produced by patients who present with various dentofacial disharmonies (DFDs), including severe overbites (Class II), underbites (Class III), and anterior open bites. Previous studies have found that patients with these malocclusion types all produce lingual sibilants and plosives with increased spectral center of gravity and increased spectral variance relative to controls. This result is puzzling since some DFD groups differ from controls in opposite ways, and it is also difficult to interpret because spectral moment measures are affected by a wide range of speech and nonspeech factors.

**Method:** To better understand the articulatory basis of these differences, we apply articulatorily interpretable spectral measures derived from multitaper spectra.

**Results:** We find that all groups of DFD patients produce /s ʃ t tʃ/ with midfrequency spectral peaks that are less prominent than those of the control group, but peak frequency measures are largely the same across all groups.

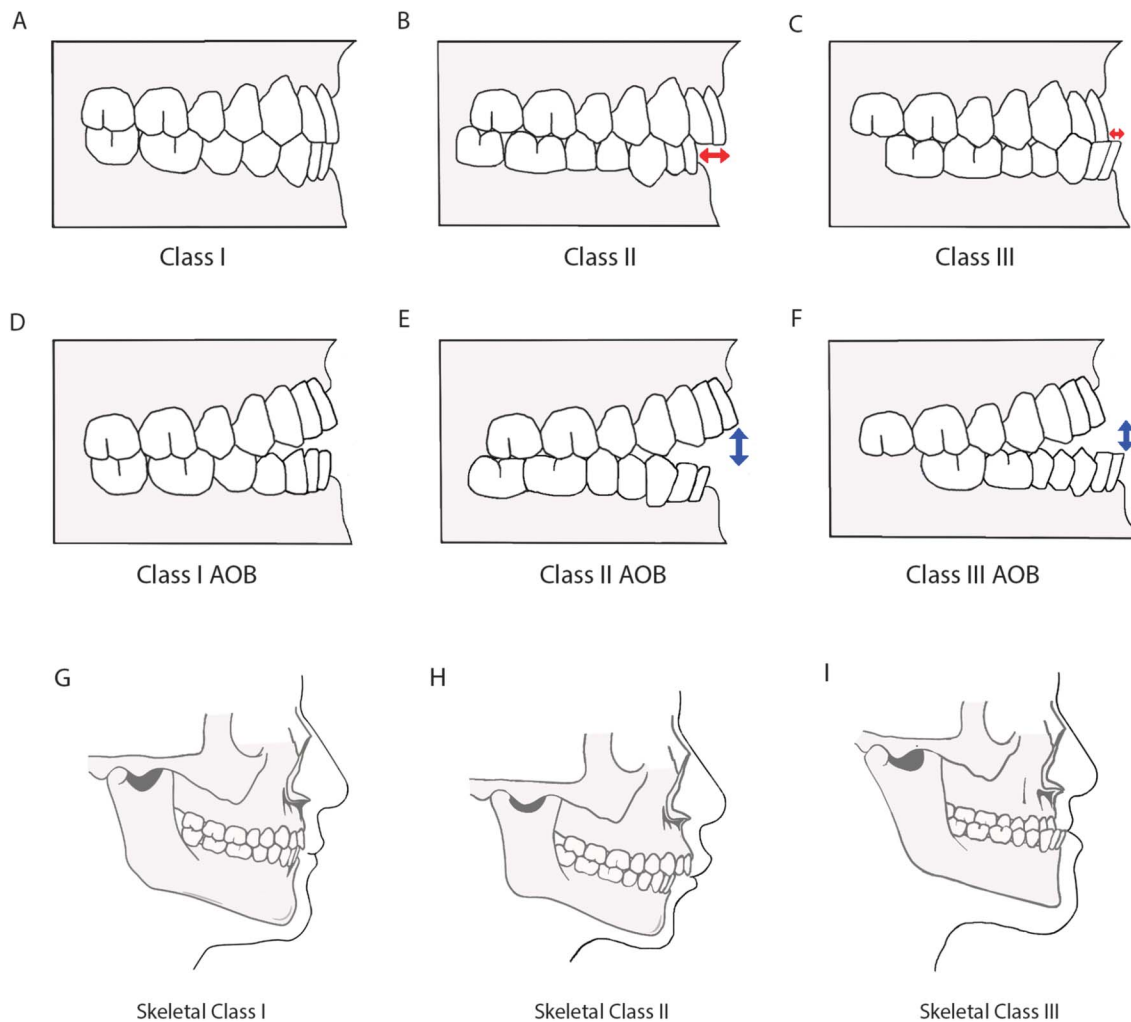
**Conclusion:** We conclude that the DFD patients differ more in sibilant noise source properties than in front cavity filter properties.

The great majority of patients who have dentofacial disharmonies (DFDs) present with speech sound disorders (Black et al., 2015; Morris et al., 2016; Ocampo-Parra et al., 2015). The passive and active articulators for speech can be aberrantly positioned in patients with DFDs; DFDs are severe skeletal disproportions requiring orthognathic jaw surgery in combination with orthodontic treatment to achieve full correction of the jaws and tooth positions. This study uses multitaper spectral analyses to quantify acoustic speech production differences between DFD population groups that reflect articulatory differences.

DFD patients present with skeletal and dental discrepancies (i.e., malocclusions) in the anterior–posterior (AP; horizontal), vertical, and/or transverse dimensions, including Class II “overbite,” Class III underbite, anterior open bite (AOB), and posterior crossbite (see Figure 1). Patients can have an isolated disproportion in one dimension or a more complex presentation with disproportions in multiple dimensions (e.g., an AP underbite in addition to a vertical AOB and a transverse posterior crossbite). Orthognathic surgery can reposition the maxilla and/or mandible to allow for ideal coupling of the teeth and jaws, improved facial esthetics, mastication, temporomandibular joint function, and speech. These surgical candidates make up 2.5% of the U.S. population and are diagnosed as having a “handicapping” DFD, which indicates abnormal facial esthetics and difficulty in chewing and speaking (Proffit et al., 1991). Among DFD patients, 90% of

Correspondence to Jeff Mielke: [jimielke@ncsu.edu](mailto:jimielke@ncsu.edu). **Disclosure:** The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.

**Figure 1.** Dental and skeletal malocclusions. (A) to (C) illustrate three different anterior–posterior relationships with a closed bite, and (D) to (F) illustrate the same three relationships combined with an anterior open bite. (G) to (I) illustrate skeletal malocclusions: (G) Class I dental and skeletal relationships with proportional, well-positioned jaws and teeth interdigitating properly. (H) Class II dental and skeletal relationships, with the mandible and mandibular teeth positioned posteriorly relative to the maxillary teeth, with excess/positive overjet. (I) Class III dental and skeletal relationships, with the mandible and mandibular teeth positioned anteriorly relative to the maxillary teeth, with deficient/negative overjet. Overjet (red arrow)—extent of horizontal (anterior–posterior) distance between the maxillary central incisors and the mandibular central incisors. It is defined as the distance from the anterior surface of the mandibular incisor to the lower (incisal) edge of the maxillary incisor. An overjet of +2 mm is normal. Overbite (blue arrow)— extent of vertical (up–down) overlap of the maxillary central incisors over the mandibular central incisors. An overlap of +2 mm is normal. No vertical overlap indicates an anterior open bite. AOB = anterior open bite.



patients with Class III underbites, 73%–87% of patients with Class II overbites, and 80% of AOB surgical patients suffer from structural speech sound distortions, as compared to 4.9% of adolescents and 3.5% of adults in the general U.S. population (Black et al., 2015; Morris et al., 2016; Ocampo-Parra et al., 2015). The large difference in the incidence of articulation errors between these populations suggests a link between jaw disproportion and articulation.

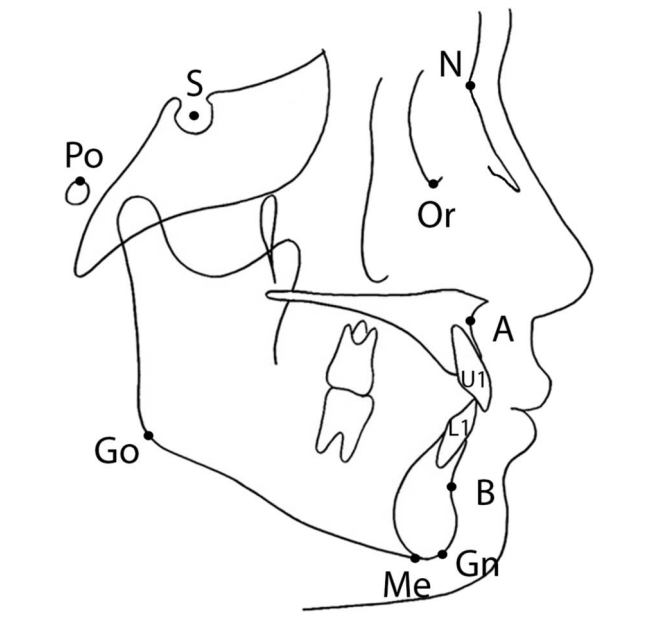
Patients can display malocclusions in the AP (front–back) or vertical positioning of their teeth, as illustrated in Figures 1A–1F. Figures 1A–1C demonstrate AP relationships,

and Figures 1D–1F demonstrate AP and vertical malocclusions. Figure 1A shows Class I occlusion—normal interdigitation of teeth in AP and vertical dimensions. Figure 1B shows Class II malocclusion—mandibular teeth are too far posterior relative to maxillary teeth with excessive/positive overjet. Figure 1C shows Class III malocclusion—mandibular teeth are too far anterior relative to maxillary teeth with deficient/negative overjet. Figure 1D illustrates Class I AOB—normal AP interdigitation of teeth, but anterior teeth fail to overlap vertically, with deficient/negative overbite. Figure 1E illustrates Class II AOB—mandibular teeth are too far posterior relative to maxillary teeth with excessive overjet combined with

anterior teeth failing to overlap, with negative overbite. Figure 1F illustrates Class III AOB—mandibular teeth are too far anterior relative to maxillary teeth with deficient overjet combined with anterior teeth failing to overlap, with negative overbite. Overjet, indicated by a red arrow, indexes the horizontal (AP) distance between the maxillary central incisors and the mandibular central incisors. It is defined as the distance from the anterior surface of the mandibular incisor to the (incisal) edge of the maxillary incisor; +2 mm is considered to be a normal overjet value. Overbite, indicated by a blue arrow, indexes the extent of vertical (up–down) overlap of the maxillary central incisors over the mandibular central incisors; +2 mm overlap is considered to be a normal overbite value. No vertical overlap indicates an AOB. Excess (> 2 mm) or deficient (< 2 mm) overjet and overbite are considered abnormal.

Skeletal and dental positions are quantified by orthodontists using lateral cephalometric radiographs (cephs); these radiographs are taken through the side of the face to generate a lateral view of the head, where right and left structures are superimposed. The ceph image is digitally traced, with labeling of key anatomical structures for linear and angular measurements of jaw and tooth positioning to quantify the degree of skeletal and dental discrepancy seen in DFD patients. In prior investigations, several ceph measurements linearly correlated with changes in spectral center of gravity and variance and, therefore, were evaluated further here. Specifically, A-point-Nasion-B-point (ANB) angle and Wits appraisal (the distance along the occlusal plane in the anterior direction from a line perpendicular to point A on the maxilla to a line perpendicular to point B on the mandible) are measures of AP or horizontal positioning of the jaws relative to one another. As ANB and Wits appraisal increase, patients have a greater Class II discrepancy, usually with a mandibular deficiency and a concomitant increase in overjet; the lower lip will be more posterior such that the mouth opening begins in a more posterior location, with a shorter front cavity (see Figures 1–2). Alternatively, when ANB and Wits appraisal decrease below normal, patients have a greater Class III underbite tendency, with a deficient maxilla and excessive mandible, such that the upper lip is more posterior and the lower lip is more anterior, with a somewhat shorter front cavity. The Frankfort-mandibular plane angle (FMA) and Sella-Nasion-Gonion-Gnathion (SN-GoGN) angle are measures of mandibular jaw angulation and vertical position; the larger these angles become, the more likely a patient has an AOB with a vertical enlargement of their anterior oral cavity (see Figure 3). The final ceph measures are the mandibular incisor to mandibular plane angle (IMPA) and the Upper-1 (central incisor) to Sella-Nasion Angle (U1-SN), which are measures of the angulation of the lower and upper central incisors, respectively; the greater the angulation, the more proclined

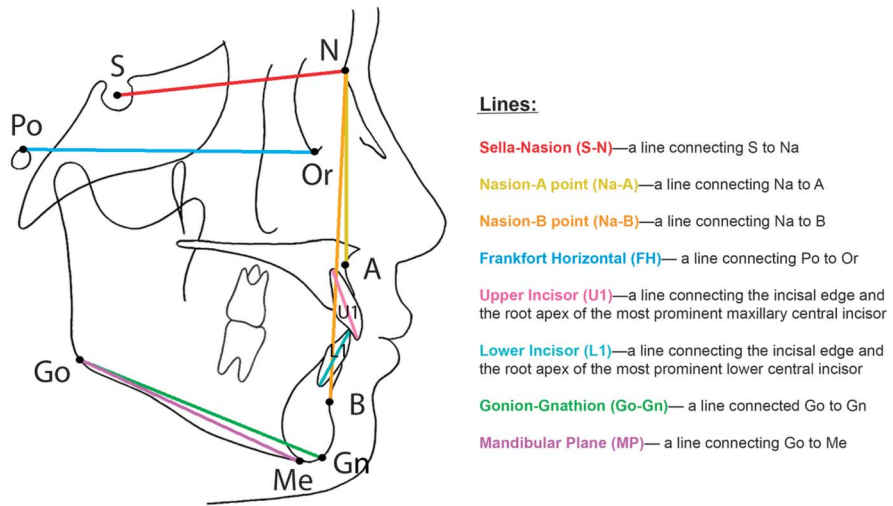
**Figure 2.** Lateral cephalogram tracing points used for linear and angular measurements. Tracing of a lateral cephalogram with landmarks labeled for linear and angular measurements. A: A point, point of the deepest concavity anteriorly on the maxillary alveolus, B: B point, point of the deepest concavity anteriorly on the mandibular symphysis, Go: anatomical gonion, a point midway between the points representing the middle of the curvature at the left and right angles of the mandible. Gn: gnathion, most outward and everted point on the profile curvature of the symphysis of the mandible, located midway between pogonion and menton. Me: menton, the lowest point on the symphysis of the mandible. N: nasion, most anterior point on the fronto-nasal suture, junction of the nasal and frontal bones at the most posterior point on the curvature of the bridge of the nose. Or: orbitale, a point midway between the lowest point on the inferior margin of the two orbits. Po: porion, the midpoint of the upper contour of the external auditory canal (Anatomic Porion). S: sella, center of the hypophyseal/pituitary fossa (sella tursica). U1: upper incisor, the most central, anterior tooth within the maxilla. L1: lower incisor, the most central, anterior tooth within the mandible.



(angled out) the front teeth are. The angles of these teeth are expected to affect the turbulence produced by a nearby constriction.

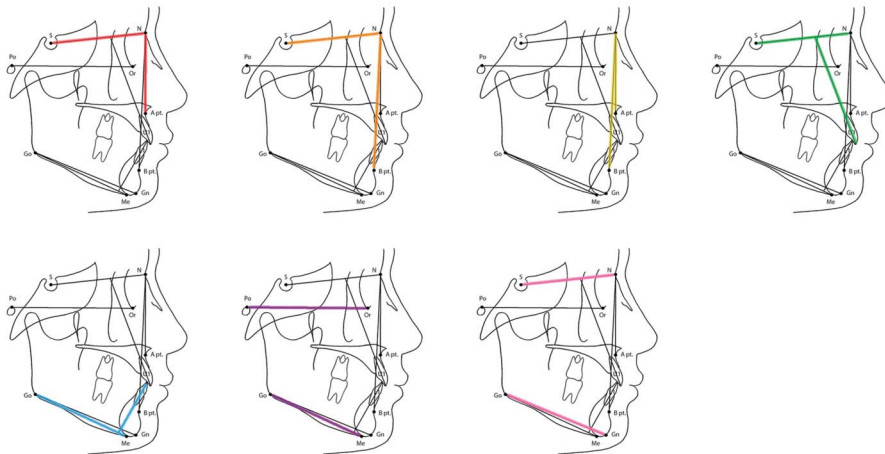
Because target-like production of obstruents in particular requires precise articulation of the tongue against passive articulators, it is unsurprising that DFD populations often present with speech sound disorders. Farronato et al. (2012) found that 50% of 6- to 10-year-old patients presenting with dental malocclusion also presented with dyslalia, with Class III malocclusion having higher dyslalia prevalence compared to Class II malocclusion. Moreover, Ocampo-Parra et al. (2015) revealed that dyslalia was present in 77.4% of students with AOB, aged 6–18 years, with distortion being the most common type of dyslalia. Amr-Rey et al. (2022) further discovered, based on auditory discrimination, a significant association between dental malocclusions (Class II, excess overjet,

**Figure 3.** Lines and angles measured from the points shown in Figure 2.



**Lines:**

- Sella-Nasion (S-N)**—a line connecting S to Na
- Nasion-A point (Na-A)**—a line connecting Na to A
- Nasion-B point (Na-B)**—a line connecting Na to B
- Frankfort Horizontal (FH)**—a line connecting Po to Or
- Upper Incisor (U1)**—a line connecting the incisal edge and the root apex of the most prominent maxillary central incisor
- Lower Incisor (L1)**—a line connecting the incisal edge and the root apex of the most prominent lower central incisor
- Gonion-Gnathion (Go-Gn)**—a line connected Go to Gn
- Mandibular Plane (MP)**—a line connecting Go to Me



**Angles:**

- SNA Angle**— Angle between S-N and Na-A lines, angle indicates the horizontal position of the maxilla relative to the cranial base using the maxillary A-point as a cephalometric landmark.
- SNB Angle**— Angle between S-N and Na-B lines, angle expresses the horizontal position of the mandible relative to the cranial base using the mandibular B-point as a cephalometric landmark.
- ANB Angle**— Angle between Na-A and Na-B lines, angle measures the relative position of the maxilla to mandible. ANB angle (ANB = A-point-Nasion-B-point) can be measured from: ANB = SNA - SNB. A positive ANB angle indicates that the maxilla is positioned anteriorly relative to the mandible (Class I or Class II malocclusion). A negative ANB angle indicates that the maxilla is positioned posteriorly relative to the mandible (Class III malocclusion).
- U1-SN**— Angle between U1 incisor line and S-N line, angular measurement determines the inclination of the central incisor relative to the anterior cranial base (larger angle is more proclined towards the lips).
- IMPA**—Mandibular Plane, also known as IMPA (Incisor Mandibular Plane Angle), defines the axial inclination between the mandibular incisor and the inferior border of the mandible. The more the incisor is labially inclined towards the lips, the greater the angle.
- FMA**— Frankfort mandibular plane angle, angle between the FH line and the MP line, the greater (higher) the angle, the more vertical the growth direction and the more complicated the orthodontic treatment.
- Sn-GoGn**— Angle between S-N line and Go-Gn line, mandibular plane can be used as a measure of mandibular growth and lower facial growth direction.

Class III, edge-to-edge bite [deficient overjet], anterior crossbite, and AOB) and sigmatisms and rhotacisms. More recent studies have confirmed the high prevalence of perceptual and quantitative distortions among DFD

patients with Class II, Class III underbites, and/or AOB. Certain consonants are more likely to be affected by different DFDs. Bilabial consonants and sibilants were found to be the most challenging for Class II patients to

produce, while Class III patients showed the most difficulty in producing stop and affricate consonants; AOB patients present with the highest prevalence of auditory distortion of fricative consonants (Jacox et al., 2022; Jhingree et al., 2022; Keyser et al., 2022; Lathrop-Marshall et al., 2022; Oliver et al., 2023). Grudziąż-Sękowska et al. (2018) and Assaf et al. (2021) found an association between malocclusion and speech distortion in children. See Bode et al. (2023) for a more extensive review of the literature.

Interestingly, recent studies that have measured the acoustic differences between malocclusion groups in production of specific consonants have found results in the opposite direction from what we might expect, given the articulatory differences between the groups. Recent studies using spectral moment analysis found increases in the first and second spectral moments (M1 = centroid frequency/center of gravity; M2 = spectral variance) of DFD cohorts including Classes II and III and AOB when compared to Class I reference subjects for the /k/, /t/, /ʃ/, /s/, and /tʃ/ sounds. 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. Linear regression models indicated that the severity of malocclusion correlated with degree of distortion for consonant sounds for all DFD cohorts: Classes II and III and AOB (Keyser et al., 2022; Lathrop-Marshall et al., 2022; Oliver et al., 2023).

Although interesting, the spectral moment results are surprising as each DFD cohort has unique jaw proportions and vocal tract anatomy, but all surgical DFD cohorts demonstrate increased first and second spectral moments (see Figure 1). The use of spectral moments may obscure differences between patient groups. Spectral moments can be readily measured from the discrete Fourier transform (DFT) of a waveform. However, because a single DFT does no averaging, it is a poor spectral estimate and the moments that result will be inaccurate (Shadle, 2006) and measures based on single DFTs arguably should not be used for fricative waveforms or any intrinsically noisy sound. See Reidy (2015) for a comparison of sibilant spectra estimated in three different ways. Spectral moment analysis is a useful method for defining the contrast between some speech sounds for specific speaker groups (Forrest et al., 1990) but has several drawbacks that can cause difficulty when interpreting the consonant production results from different DFD cohorts. As discussed by Koenig et al. (2013) and Shadle (2023), spectral moments are sensitive to multiple articulatory movements and environmental noise sources. This may be

obscuring differences across DFD cohorts. Spectral moment analysis offers a global look at consonant spectra, but cannot provide detailed, unambiguous assessments of spectral features needed to determine why contrasting oral anatomy, on opposite ends of the AP and vertical dimensions, are resulting in similar spectral outcomes.

### **Multitaper Spectral Analysis of DFD Speech**

To understand the reasons for these observed differences in spectral moments, it is necessary to apply acoustic measurements that have more direct articulatory interpretations. This includes measures of spectral peak frequency and amplitude. Since conventional fast Fourier transform-based approaches to spectral estimates are noisy (due to the lack of any averaging), they make it challenging to produce reliable spectral peak measurements. Measuring the spectral peaks, frequencies, and amplitudes with high temporal precision is more practical with single multitaper spectra. Acoustic measurements such as peak frequency and spectral amplitude differences measured using multitaper spectra more closely reflect articulatory changes, such as labiality (which lowers midfrequency peaks), changes in the size of the front cavity over time, and aerodynamic effects of the decrease in the oral constriction over time (Jesus & Shadle, 2002; Koenig et al., 2013; Shadle, 2023).

This study applies multitaper spectral analysis to speech recordings from presurgical DFD populations to evaluate how the severity of Classes II and III and AOB malocclusions correlate with speech distortion of consonants. Specifically, peak frequency and spectral amplitude difference are measured for five consonants /s t ʃ tʃ k/. Peak frequency is measured as the overall spectral peak and is expected to be inversely correlated with the size of the front cavity between the tongue constriction and the mouth opening. We note that the resonant frequency of the front cavity is more closely related to a smaller peak that is lower in frequency and amplitude than the overall spectral peak (Shadle, 2023), but a procedure for consistently locating that peak has not been described. Accordingly, we have chosen to focus on the main peak. Spectral amplitude difference is measured as the difference between the midfrequency spectral peak and the low-frequency spectral trough and is expected to be associated with the strength of the air turbulence at the noise source and the antiresonance generated by the decoupling action of the small constriction.

On the horizontal plane, Class II patients differ from Class I controls in that overjet tends to be greater than 2 mm, ANB angle is greater than 4°, and Wits appraisal is greater than 3 mm. As overjet, ANB angle, and Wits appraisal increase, the size of the front cavity

for the consonants /s t ʃ tʃ k/ will likely decrease. Thus, we hypothesize that peak frequency will be higher for Class II patients and will increase as overjet, ANB angle, and Wits appraisal increase. Because Class II patients may not position the tongue against the alveolar ridge and postalveolar region with as much precision as Class I speakers, making a weaker constriction for stops and fricatives, we hypothesize that spectral amplitude difference for the consonants /s t ʃ tʃ/ will be lower for Class II patients. /k/ is not anticipated to have this same effect for amplitude difference, given the constriction location at the velum, which will largely be unaffected by overjet, ANB angle, and Wits appraisal deviations.

Class III patients differ from Class I controls on the horizontal plane in that overjet tends to be less than 2 mm, ANB angle is less than 0°, and Wits appraisal is less than 0 mm. Similar to Class II patients, front cavity size is likely to decrease for Class III patients compared to Class I controls, meaning that peak frequency measures will likely increase in consonant production for Class III patients. Again, because Class III patients likely make less precise constrictions for stops and fricatives than Class I controls, amplitude difference for these consonants is expected to be lower for Class III patients than Class I controls. Again, the amplitude difference for /k/ will likely not differ between Class I and Class III patients due to its posterior constriction location.

Patients may also have malocclusions that deviate from control populations on the vertical dimension. Patients with open bites have negative overbite values, high SN-GoGN values, and high FMA angles (see Figures 2–3 for more descriptions of these values). This malocclusion is likely to interact with a speaker's lip opening and front cavity size, which will affect peak frequency measures. As open bites become more severe, patients may have a larger front cavity, which will decrease the overall peak frequency. This effect is likely to depend on consonant quality. English /ʃ tʃ/ are produced with rounded lips, which patients may use to a greater degree to reach target-like peak frequency values. For consonants typically produced with a lip opening narrow enough that the front cavity approximates a Helmholtz resonator, enlargement of the opening is expected to increase the front cavity resonant frequency. For amplitude difference measures, patients with open bites may not be able to form as tight of a constriction between the tongue tip or blade and teeth, which will overall decrease amplitude difference values for this population. However, patients with an open bite are not anticipated to differ from control group speakers in terms of amplitude difference for the production of /k/. This is again because of the posterior constriction location of /k/, which is likely to be unaffected by overbite, SN-GoGN, or FMA angles.

## Method

### Participants

Two hundred eighty-four people participated in the study, including 223 DFD patients and 61 reference controls with Class I occlusion and skeletal base. The DFD patients, representing five malocclusion types, were consecutively enrolled from a DFD clinic at the University of North Carolina (UNC) Adams School of Dentistry (ASOD): Class II closed bite ( $n = 46$ ), Class III closed bite ( $n = 109$ ), Class I AOB ( $n = 7$ ), Class II AOB ( $n = 17$ ), and Class III AOB ( $n = 44$ ) malocclusions. Subsets of the recordings were previously analyzed using spectral moment analysis in Keyser et al.'s (2022) study of speech with AOB, Lathrop-Marshall et al.'s (2022) study of speech with Class III malocclusion and Oliver et al.'s (2023) study of speech with Class II malocclusion. All patients were referred for a surgical evaluation and screened by a board-certified orthodontist prior to enrollment based on their malocclusions (Lathrop-Marshall et al., 2022). All DFD patients were 12–53 years old ( $M = 20.7$ ); 125 were female and 98 were male. The participants in the control group were 15–38 years old ( $M = 22.3$ ), with 39 females and 22 males. Orthodontic and surgical records were collected and consisted of occlusal measurements, dental models, photos (intraoral and extraoral), and panoramic and cephalogram radiographs, as described in Lathrop-Marshall et al. (2022). Ethics approval was granted by the institutional review board of UNC ASOD (#18–1406 & #19–1196). More detailed information is included in Table 1.

### Materials

Twenty English target words containing word-initial /s t ʃ tʃ k/ before the vowels /i u æ a/ were embedded in the carrier phrase, “Say \_\_\_ again.” as shown in Table 2.

### Procedures

Phrases were randomized and presented one at a time on a computer screen in three different random sequences. 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 unidirectional condenser microphone (AKG Pro Audio C520 Professional) and digitized at a sample rate of 44.1 kHz with 16 bits per sample using a CSL module.

### Analysis

The previous spectral moment analysis relied on manual selection of consonant measurement points and

**Table 1.** Demographic information for the six groups of participants.

Demographics	Class III CB	Class III AOB	Class II CB	Class II AOB	Class I AOB	Control
Age in years ( <i>M</i> )	20.3 ( <i>n</i> = 109)	21.1 ( <i>n</i> = 44)	20.4 ( <i>n</i> = 46)	22.2 ( <i>n</i> = 17)	20.7 ( <i>n</i> = 7)	22.2 ( <i>n</i> = 61)
Age in years (range)	12–53	14–40	14–38	13–35	15–33	15–38
Gender	50.5% Female ( <i>n</i> = 55)	65.9% Female ( <i>n</i> = 29)	56.5% Female ( <i>n</i> = 26)	58.8% Female ( <i>n</i> = 10)	71.4% Female ( <i>n</i> = 5)	63.9% Female ( <i>n</i> = 39)
	49.5% Male ( <i>n</i> = 54)	34.1% Male ( <i>n</i> = 15)	43.5% Male ( <i>n</i> = 20)	41.2% Male ( <i>n</i> = 7)	28.6% Male ( <i>n</i> = 2)	36.1% Male ( <i>n</i> = 22)
Race	35.8% African American ( <i>n</i> = 39)	36.4% African American ( <i>n</i> = 16)	4.3% African American ( <i>n</i> = 2)	29.4% African American ( <i>n</i> = 5)	57.1% African American ( <i>n</i> = 4)	14.75% African American ( <i>n</i> = 9)
	8.3% Asian/Pacific Islander ( <i>n</i> = 9)	9.1% Asian/Pacific Islander ( <i>n</i> = 4)	2.2% Asian/Pacific Islander ( <i>n</i> = 1)	5.9% Asian/Pacific Islander ( <i>n</i> = 1)	0% Asian/Pacific Islander ( <i>n</i> = 0)	16.4% Asian/Pacific Islander ( <i>n</i> = 10)
	52.3% Caucasian ( <i>n</i> = 57)	50.0% Caucasian ( <i>n</i> = 22)	93.5% Caucasian ( <i>n</i> = 43)	64.7% Caucasian ( <i>n</i> = 11)	42.9% Caucasian ( <i>n</i> = 3)	68.85% Caucasian ( <i>n</i> = 42)
	1.8% Native American ( <i>n</i> = 2)	0% Native American ( <i>n</i> = 0)	0% Native American ( <i>n</i> = 0)	0% Native American ( <i>n</i> = 0)	0% Native American ( <i>n</i> = 0)	0% Native American ( <i>n</i> = 0)
	1.8% Other ( <i>n</i> = 2)	4.5% Other ( <i>n</i> = 2)	0% Other ( <i>n</i> = 0)	0% Other ( <i>n</i> = 0)	0% Other ( <i>n</i> = 0)	0% Other ( <i>n</i> = 0)
Ethnicity	12.8% Hispanic ( <i>n</i> = 14)	11.4% Hispanic ( <i>n</i> = 5)	6.5% Hispanic ( <i>n</i> = 3)	0% Hispanic ( <i>n</i> = 0)	0% Hispanic ( <i>n</i> = 0)	9.8% Hispanic ( <i>n</i> = 6)
	87.2% non-Hispanic ( <i>n</i> = 95)	88.6% non-Hispanic ( <i>n</i> = 39)	93.5% non-Hispanic ( <i>n</i> = 43)	100% non-Hispanic ( <i>n</i> = 17)	100% non-Hispanic ( <i>n</i> = 7)	90.2% non-Hispanic ( <i>n</i> = 55)
Bonded labial appliances (currently in braces)	45.9% ( <i>n</i> = 50)	63.6% ( <i>n</i> = 28)	56.5% ( <i>n</i> = 26)	35.3% ( <i>n</i> = 6)	42.9% ( <i>n</i> = 3)	16.4% ( <i>n</i> = 10)
Overjet mean, range (in mm)	−3.7, −13 to 3	−3.3, −17 to 2	7.9, 4 to 18	6.6, 1 to 15	3.1, 0 to 11	2.3, 1 to 5.5
Overbite mean, range (in mm)	1.9, 0 to 6	−3.0, −9 to 0	4.7, 0 to 12	−3.9, −12 to 1	−6.1, −1 to −10	2.3, 1 to 4
Ceph mean, ranges:						
ANB mean, range (in degrees)	−3.7 (range: −12.0 to 6.8)	−1.8 (range: −12.8 to 5.9)	5.6 (range: −0.3 to 12.9)	5.6 (range: 1.8 to 11.5)	4.0 (range: 2.2 to 7.2)	N/A
IMPA (in degrees)	84.4 (range: 61.5 to 102.9)	87.8 (range: 62.8 to 104.5)	94.6 (range: 75.8 to 116.1)	91.4 (range: 79.2 to 114.4)	93.6 (range: 72.1 to 101.5)	N/A
Wits (mm)	−10.4 (range: −23.0 to 4.1)	−9.9 (range: −21.6 to −1.7)	6.0 (range: −1.6 to 14.8)	1.6 (range: −8.6 to 12.1)	−0.1 (range: −3.5 to 4.6)	N/A
FMA (in degrees)	23.4 (range: 8.8 to 49.0)	28.0 (range: 13.5 to 64.6)	23.6 (range: 2.5 to 45.3)	32.4 (range: 20.6 to 45.8)	27.0 (range: 16.6 to 34.0)	N/A
SN-GoGn (in degrees)	29.9 (range: 17.8 to 47.6)	34.1 (range: 22.1 to 51.3)	31.4 (range: 13.0 to 52.8)	41.0 (range: 30.6 to 51.2)	34.6 (range: 27.3 to 44.3)	N/A

*Note.* CB = closed bite; AOB = anterior open bite; ANB = A-point-Nasion-B-point; IMPA = mandibular incisor to mandibular plane angle; FMA = Frankfort-mandibular plane angle; SN-GoGn = Sella-Nasion-Gonion-Gnathion; N/A = not applicable, data not available.

**Table 2.** Stimuli.

	/s/	/t/	/ʃ/	/tʃ/	/k/
/i/	see	tea	she	cheap	key
/u/	sue	too	shoe	chew	coo
/æ/	sack	tap	shack	chap	cap
/ɑ/	sock	top	shock	chop	cop

extraction of spectral moment values from DFT spectra using TF32 software (Natour & Saleem, 2009), which applied pre-emphasis. The current analysis differed in several ways. We used forced alignment and acoustic measures to automatically identify measurement points, and we created multitaper spectra without pre-emphasis. All other methodological choices remained the same as in previous studies.

The recordings were made over the course of multiple years, and they were processed in two batches. All target phrases were typically produced in the same order in all recordings. However, not all of them can be force-aligned using a transcript based directly on the prompts. This is because some of the recordings contain extraneous speech and some target phrases were repeated, misread, or produced out of order. We created a transcript for each recording based on the prompt list, and manually adjusted it as necessary for each recording. For the second batch of recordings, we directly created a custom transcript for each recording using Pyannote (Bredin et al., 2020) and Vosk (Vosk Speech Recognition Toolkit, 2023) and then made corrections as needed to match the speech in the recordings. Both techniques resulted in accurate transcripts, which were then used as the input to forced alignment. The Montreal Forced Aligner (McAuliffe et al., 2017) was used to segment and align the recordings using english\_us\_arpa acoustic models. Segmentation was hand-corrected for all recordings.

The fricatives /s ʃ/ were measured in a 20-ms window centered on the midpoint of the fricative interval, which can be located based on the segmentation produced by forced alignment. The other consonants /t tʃ k/ were measured using a 20-ms window whose left edge coincided with the consonant's release. To make these measurements, it was necessary to first locate the consonant's release. We created a segmentation scheme following Cronenberg et al. (2020), who used the relative values of the voicing probability measure from Pitch Estimation Filter with Amplitude Compression (PEFAC; Gonzalez & Brookes, 2014) and their own high-frequency energy signal to identify pre- and postaspiration in Andalusian Spanish. Following Cronenberg et al. (2020), we created a high-frequency signal by first applying preemphasis twice to effectively boost the signal 12 dB per octave, then high-pass filtering at 3 kHz using the wrassp package (a

wrapper for libassp; Winkelmann & Raess, 2014) and calculating the RMS intensity of this high-frequency sound signal. Note that these filters are part of the segmentation procedure but not part of the measurement procedure, which used no pre-emphasis. Since we are using this signal specifically for segmentation, we diverged from Cronenberg et al. (2020) by using a time step of 1 ms for the intensity signal and not applying a Butterworth filter.

To use these signals for segmentation, we found the time point of the maximum high-frequency signal value (within the consonant release/noise interval) and the preceding minimum value (within the closure). Between these two time points, the time of the steepest rise of the high-frequency signal was taken as the start of the release and therefore the left edge of the 20-ms measurement window. Multitaper spectra were measured from the designated 20-ms windows in each consonant using the spectRum package (Reidy, 2013) for R (R Core Team, 2000), with eight tapers and a bandwidth parameter of 4. Pre-emphasis was not applied for the purposes of making spectral measurements.

This study focuses on two acoustic measurements, which are designed to access different articulatory parameters, following Jesus and Shadle (2002), Koenig et al. (2013), and Shadle (2023). The frequency of the main spectral peak (measured in the 1–22 kHz range, which is effectively 1–20 kHz due to the frequency response of the microphone) is expected to be associated with the size of the front cavity between the lingual constriction and the mouth opening. The midfrequency spectral peak (Jesus & Shadle, 2002; Koenig et al., 2013; Shadle, 2023) is more directly related to the front cavity resonance than the main peak. However, there is not a well-defined method for locating the midfrequency spectral peak in the bursts of unrounded consonants such as /s/. In light of this difficulty, we measure the main spectral peak, which generally coincides with the midfrequency peak for the rounded postalveolar consonants /ʃ tʃ/. For /s t/, it is sometimes the midfrequency peak and sometimes a higher frequency peak. For /k/, this peak is typically close to the F2 frequency of the following vowel.

Spectral amplitude difference (the difference in amplitude between the midfrequency spectral peak and a lower frequency spectral trough measured between 1 kHz and the main spectral peak) is expected to be associated with the noise source (and the antiresonance caused by the narrow constriction). /k/ lacks a midfrequency spectral peak, so it was not measured in this way. The spectral peak and amplitude difference measurements for two sample tokens are illustrated in Figure 4.

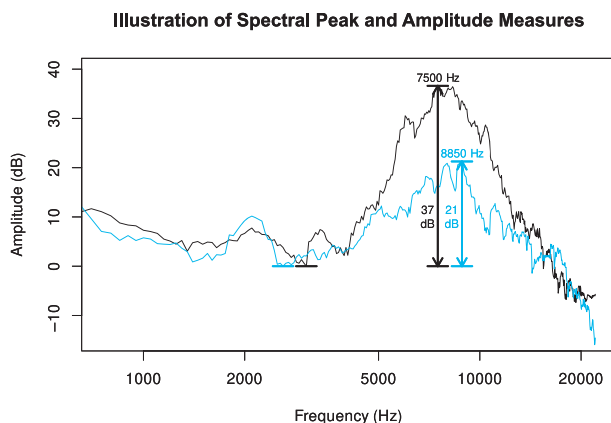
We applied two methods for excluding erroneous measurements. We plotted all tokens by peak frequency and peak amplitude difference and identified obvious outliers,



which had unreasonably low peak frequencies (< 2200 Hz for /s/, < 1600 Hz for /t tʃ/, < 1400 Hz for /ʃ/, and < 1200 Hz for /k/). We also listened to sample sound clips from all recordings and counted how many included excess airflow noise due to suboptimal microphone placement. The recordings with excess airflow noise accounted for many of the unreasonable peak frequency measurements. Twenty-five participants not listed above (22 DFD patients and three controls) were excluded due to noisy recordings, related to suboptimal microphone placement. The other tokens with unreasonable measurements ( $n = 908$ ) were excluded, leaving 16,167 tokens for analysis. No other outlier removal procedures were applied.

To evaluate the relevance of these acoustic measures for perceptual judgments of speech distortion, we used a set of perceptual judgments collected by Jhingree et al. (2022), which included the consonants /s f t k/ for 25 of the participants whose acoustic data are included in this study (10 Class III closed bite, four Class III AOB, one Class I AOB, and eight Class I controls). The listeners were 30 speech specialists who were asked to rate the level of perceived distortion in the initial consonant of words presented in isolation on a visual analog scale ranging from 0 (completely clear) to 100 (severely distorted). The perceptual responses and the acoustic measures were averaged to produce a single value of each acoustic and perceptual measure for each of the 100 speaker-consonant combinations. A linear regression was performed for each of the eight consonant-acoustic measure combinations.

**Figure 4.** Spectral peak and amplitude difference measurements for a token of /s/ before /i/ in the word “see,” produced by a female control speaker (in black) and produced by a female Class III patient (in blue). The spectral peaks and troughs are indicated by horizontal line segments. The amplitude difference of each main spectral peak (relative to the lower spectral trough) is indicated by the height of the arrow. The peak’s frequency is indicated by the arrow’s horizontal position.



Mixed-effects linear regression models were run for peak amplitude difference (dB) and peak frequency (log Hz) as the dependent variables, with the formula shown in Equation 1 using the lmerTest package (Kuznetsova et al., 2017). Patient group and gender are included as fixed-effects, and word and speaker are included as random effects. We use .05 as our  $\alpha$  criterion. Marginal and conditional  $R^2$  values were calculated using the sjPlot package (Lüdtke, 2023), using the method described by Nakagawa et al. (2017).

$$\text{lmer}(\text{acoustic\_measure} \sim \text{class} \times \text{AOB} \times \text{gender} \\ \times \text{consonant} + \text{vowel} + (1 | \text{word}) + (1 | \text{speaker})) \quad (1)$$

As described above, several cephalometric measures are associated with each patient: three vertical measures (overbite, SN-GoGn, and FMA), three AP measures (overjet, ANB, and Wits appraisal), and two incisor angulation measures (U1-SN, IMPA). Clearly, some of these measures are correlated with one another. A further complication is that all of these cephalometric measures exist for patients in the malocclusion groups, but for the control participants, we generally have only overbite and overjet measures. We conducted random forest analyses (using the package Random Forest based on Breiman, 2001) including all of the available cephalometric measures (but only the 220 patients for whom we have these measures).

Since the random forest analyses show the relative importance of variables but not the direction of their effects, we also made two mixed-effects linear regressions for each cephalometric measure, with peak amplitude difference and peak frequency again as the dependent measures, using the formula in Equation 2.

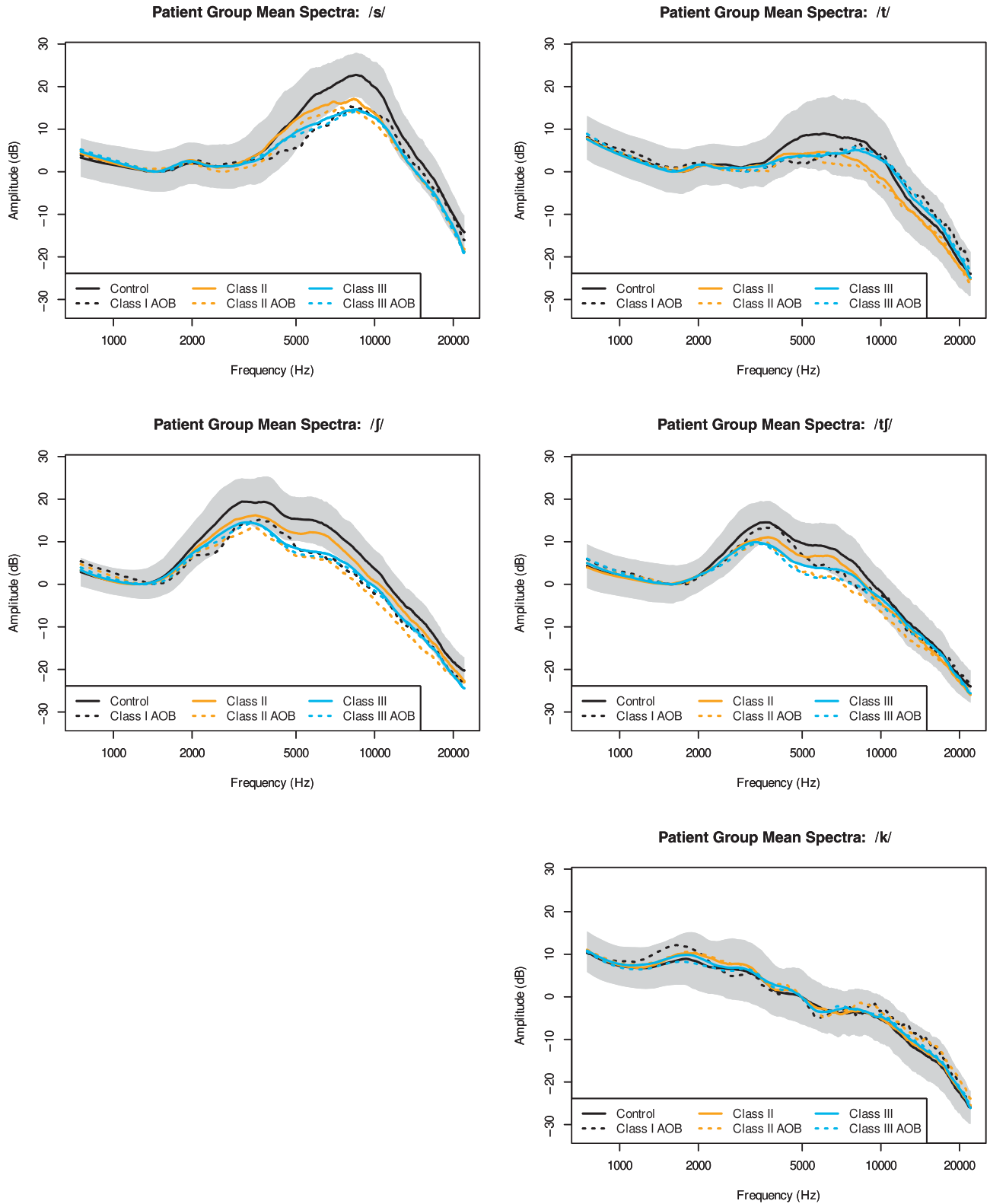
$$\text{lmer}(\text{acoustic\_measure} \sim \text{ceph\_measure} + \text{gender} \\ + (1 | \text{consonant}) + (1 | \text{word}) + (1 | \text{speaker})) \quad (2)$$

## Results

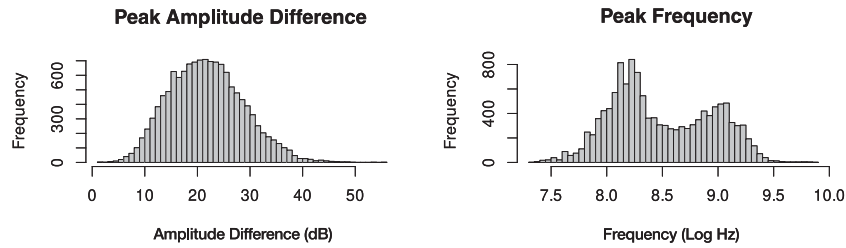
The mean spectra for each consonant for each patient group are shown in Figure 5. The spectra are adjusted so that 0 dB corresponds to the low-frequency trough. For /s f t tʃ/, it is clear that the control group has more prominent spectral peaks, and that the frequencies of these peaks appear similar across groups. The /k/ spectra appear similar across patient groups. /k/ varies more across vowel contexts than any of the other consonants.

The distributions of the spectral peak measurements for all included tokens (all speakers and all consonant–vowel combinations) are shown in Figure 6. Categorical differences between consonants are visible in the histogram for peak frequency but not in the histogram for peak amplitude difference.

**Figure 5.** Mean spectra averaged for each patient group for each consonant. AOB = anterior open bite.



**Figure 6.** Distributions of measurements for the two dependent measures included in the regression models.



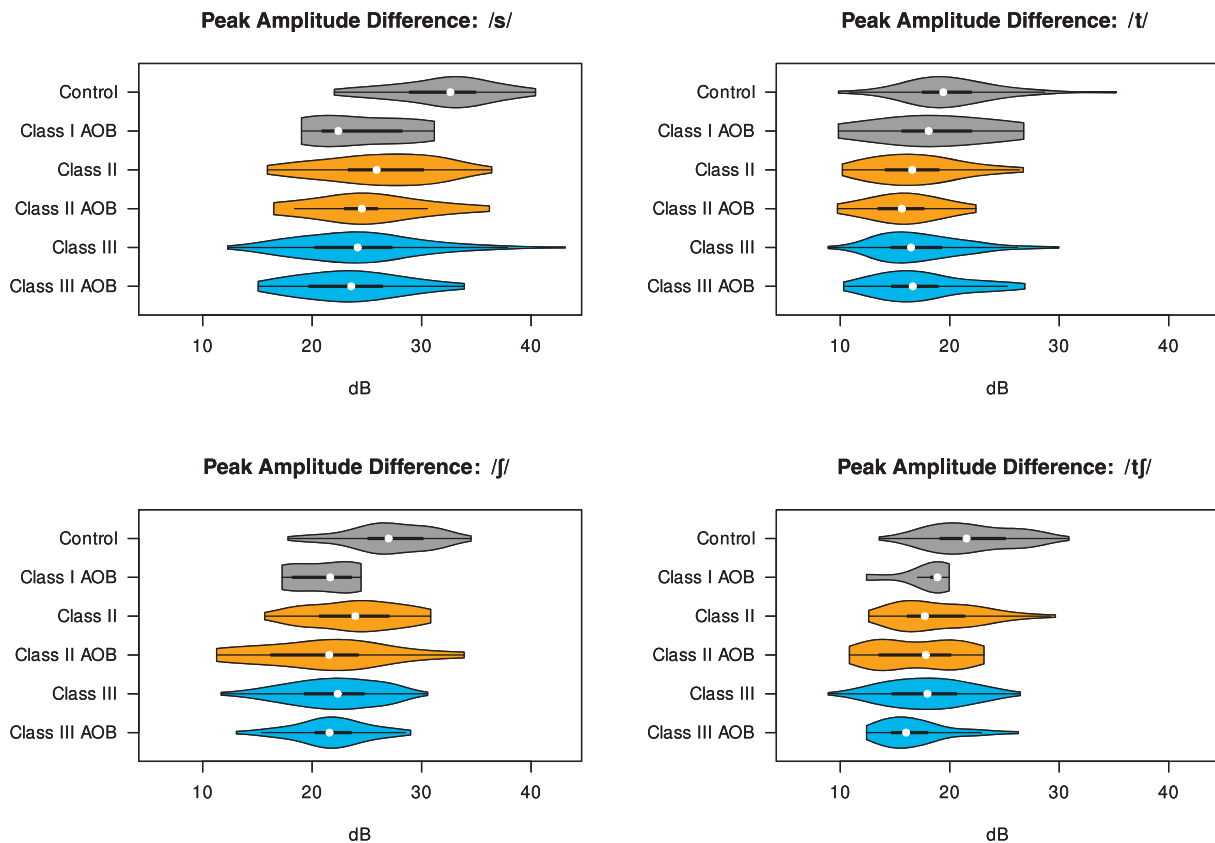
### Spectral Differences Between Groups

The most salient difference between mean spectra in Figure 5 is the difference in spectral peak amplitude between the control group and the other groups. Spectral amplitude difference (the difference in amplitude between the peak and the lower frequency trough) is expected to relate to the strength of the noise source. This measure was compared across patient groups for consonants except /k/, which shows no apparent differences in spectral peak frequency or amplitude. Figure 7 shows separate violin plots for each of the four consonants for the six groups of

participants. Table 3 shows the results of the regression with spectral amplitude difference as the dependent variable.

The Class II group has significantly smaller peak amplitude differences than the Class I controls (Estimate =  $-2.56$ ,  $SE = 0.77$ ,  $t = -3.31$ ,  $p = .001$ ) and so does the Class III group (Estimate =  $-1.86$ ,  $SE = 0.63$ ,  $t = -2.96$ ,  $p = .003$ ). There is no main effect for AOB ( $p = .0776$ ), which does not seem to affect the base level consonant /t/, but it is associated with significantly smaller peak amplitude differences for all three other consonants /s/ (Estimate =  $-7.20$ ,  $SE = 1.03$ ,  $t = -7.01$ ,  $p < .001$ ), /j/ (Estimate =

**Figure 7.** Peak amplitude difference (dB) for each consonant across participant groups. AOB = anterior open bite.



**Table 3.** Results of linear mixed-effects regression involving participant group differences in peak amplitude difference. Reference levels are Class I, closed bite, female, /t/, and following /a/.

Predictors	Estimates	SE	Statistic	p
Intercept	20.20	0.54	37.70	< .001
Class II	-2.56	0.77	-3.31	.001
Class III	-1.86	0.63	-2.96	.003
AOB	0.41	1.44	0.29	.776
Gender [M]	-2.19	0.80	-2.74	.006
Consonant /s/	11.92	0.41	28.73	< .001
Consonant /j/	7.07	0.42	17.03	< .001
Consonant /tj/	1.38	0.42	3.32	.001
Following [æ]	0.10	0.27	0.37	.710
Following [i]	0.92	0.27	3.38	.001
following [u]	1.21	0.27	4.45	< .001
Class II × AOB	-2.91	1.85	-1.57	.116
Class III × AOB	-0.54	1.60	-0.34	.734
Class II × gender [M]	-0.33	1.21	-0.28	.783
Class III × gender [M]	-0.92	0.98	-0.94	.349
AOB × gender [M]	-4.40	2.69	-1.64	.101
Class II × consonant /s/	-1.94	0.55	-3.50	< .001
Class III × consonant /s/	-4.82	0.44	-10.93	< .001
Class II × consonant /j/	-0.10	0.55	-0.19	.851
Class III × consonant /j/	-3.27	0.44	-7.39	< .001
Class II × consonant /tj/	-0.27	0.56	-0.48	.633
Class III × consonant /tj/	-1.83	0.44	-4.11	< .001
AOB × consonant /s/	-7.20	1.03	-7.01	< .001
AOB × consonant /j/	-7.38	1.03	-7.19	< .001
AOB × consonant /tj/	-4.74	1.05	-4.53	< .001
Gender [M] × consonant /s/	0.11	0.56	0.20	.845
Gender [M] × consonant /j/	0.16	0.56	0.29	.775
Gender [M] × consonant /tj/	1.88	0.56	3.33	.001
(Class II × AOB) × gender [M]	6.34	3.22	1.97	.049
(Class III × AOB) × gender [M]	3.70	2.91	1.27	.204
(Class II × AOB) × consonant /s/	7.74	1.33	5.81	< .001
(Class III × AOB) × consonant /s/	6.00	1.14	5.28	< .001
(Class II × AOB) × consonant /j/	7.74	1.33	5.81	< .001
(Class III × AOB) × consonant /j/	6.35	1.14	5.58	< .001
(Class II × AOB) × consonant /tj/	5.65	1.35	4.18	< .001
(Class III × AOB) × consonant /tj/	2.78	1.16	2.41	.016
(Class II × gender [M]) × consonant /s/	-2.13	0.86	-2.48	.013
(Class III × gender [M]) × consonant /s/	-0.43	0.70	-0.61	.539
(Class II × gender [M]) × consonant /j/	-0.37	0.86	-0.43	.667
(Class III × gender [M]) × consonant /j/	1.30	0.69	1.87	.062
(Class II × gender [M]) × consonant /tj/	-0.25	0.86	-0.29	.769
(Class III × gender [M]) × consonant /tj/	0.24	0.70	0.35	.728
(AOB × gender [M]) × consonant /s/	1.31	1.93	0.68	.499
(AOB × gender [M]) × consonant /j/	6.92	1.93	3.58	< .001
(AOB × gender [M]) × consonant /tj/	5.17	1.94	2.66	.008
(Class II × AOB × gender [M]) × consonant /s/	-2.10	2.33	-0.90	.368
(Class III × AOB × gender [M]) × consonant /s/	-0.50	2.09	-0.24	.810
(Class II × AOB × gender [M]) × consonant /j/	-10.77	2.32	-4.63	< .001
(Class III × AOB × gender [M]) × consonant /j/	-4.00	2.09	-1.91	.056
(Class II × AOB × gender [M]) × consonant /tj/	-7.75	2.34	-3.32	.001
(Class III × AOB × gender [M]) × consonant /tj/	-2.68	2.10	-1.27	.203

(table continues)

**Table 3.** (Continued).

Predictors	Estimates	SE	Statistic	<i>p</i>
Random effects				
$\sigma^2$			25.21	
$\tau_{00}$ speaker			6.61	
$\tau_{00}$ word			0.12	
ICC			0.21	
N word			16	
N speaker			284	
Observations			12,994	
Marginal $R^2$ /conditional $R^2$			.360/.495	

Note. Values in bold are significant by  $p < .05$  convention. SE = standard error; AOB = anterior open bite; ICC = intraclass correlation coefficient.

-7.38,  $SE = 1.03$ ,  $t = -7.19$ ,  $p < .001$ ), and /t/ (Estimate = -4.74,  $SE = 1.05$ ,  $t = -4.53$ ,  $p < .001$ ). These three consonants also have even smaller peak amplitude differences (relative to /t/) for Class III patients, as does /s/ for Class II patients. These results can be observed in the violin plots: AOB means are consistently lower than the corresponding closed bite groups for /s f tʃ/ but not /t/. Several of the interactions between Class II/III and AOB shown in Table 3 are significantly positive, indicating that the combined effects of AP and open bite disproportions on turbulence noise are less than additive.

Table 3 also shows some expected phonetic differences: the fricatives and affricate /s f tʃ/ have greater peak amplitude differences than the stop /t/, and consonants are noisier before high vowels /i u/. Males have smaller peak amplitude differences than females (Estimate = -2.19,  $SE = 0.80$ ,  $t = -2.74$ ,  $p = .006$ ), but not for /tʃ/. There are a few three- and four-way interactions involving gender that we did not predict and we do not have interpretations for.

The frequency of the main spectral peak is expected to correlate somewhat with front cavity size (although not as closely as the midfrequency peak, cf. Shadle, 2023). This measure was compared across patient groups for consonants except /k/, which shows no apparent differences in spectral peak frequency or amplitude. Figure 8 shows separate violin plots for each of the four consonants for the six groups of participants. The peak frequency does not vary as noticeably between groups as the amplitude does.

The differences in peak frequency are small. Relative to Class I controls, Class III patients have significantly higher peak frequency than Class I controls (Estimate = 0.14,  $SE = 0.03$ ,  $t = 4.14$ ,  $p < .001$ ), although this main effect is mostly offset by significant opposite interactions between Class III and other consonants, as shown in Table 4. Similarly, the AOB group has significantly higher peak frequency than Class I controls (Estimate = 0.19,

$SE = 0.08$ ,  $t = 2.40$ ,  $p = .016$ ), and this is more than offset by significant opposite interactions between AOB and other consonants. We find some unsurprising phonetic effects: /s/ has much higher peak frequency than /t/, and the postalveolar consonants /ʃ tʃ/ have much lower peak frequency, and peak frequency is higher before front vowels. Peak frequency is lower for males, particularly for the fricatives /s ʃ/, and it is not obvious whether this is socially or physiologically motivated.

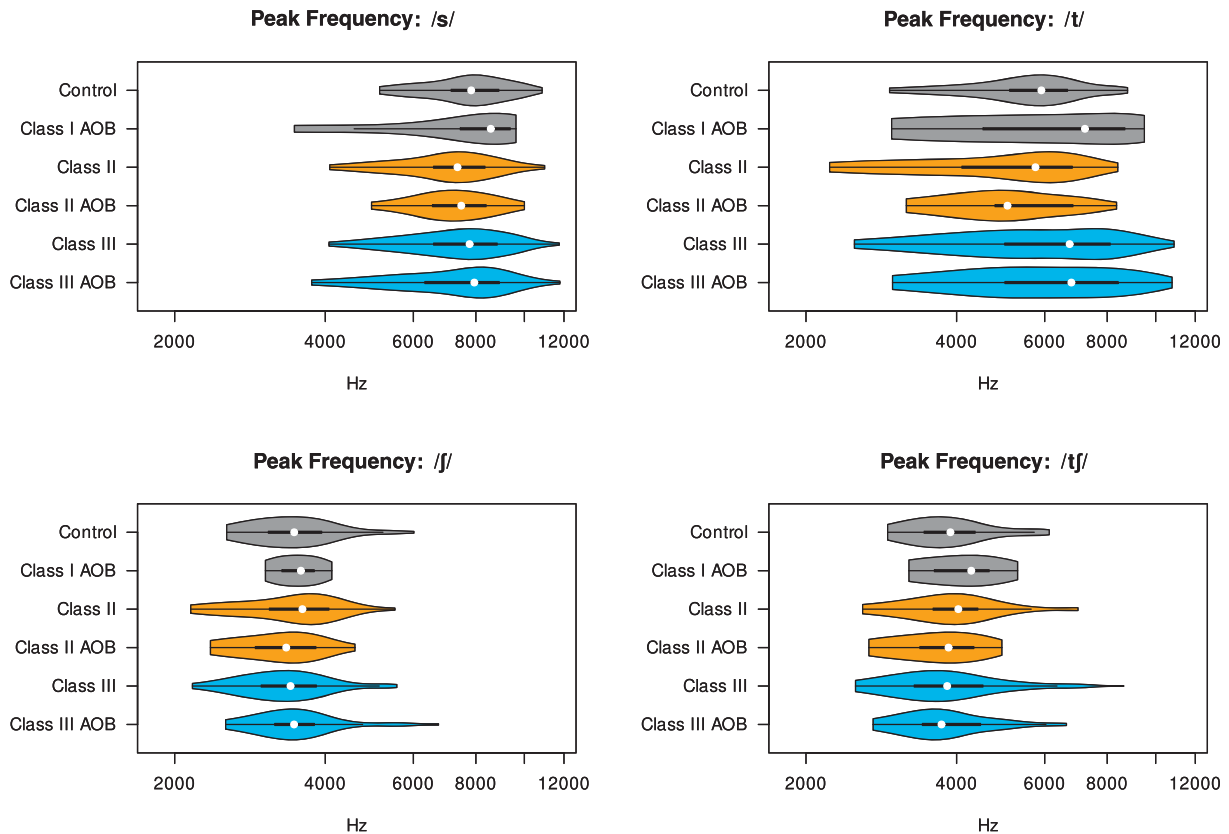
### Comparison of Acoustic Measures With Perceptual Judgments

The two acoustic measures (peak amplitude difference and peak frequency) are compared with perceptual measures of consonant distortion in Figure 9, and the results of the eight linear regressions are shown in Table 5. The relationship between the perceptual rating and peak amplitude difference is significant for /s/ (Estimate = -1.2452,  $SE = 0.1947$ ,  $t = -6.3950$ ,  $p < .0001$ ) and /ʃ/ (Estimate = -1.2452,  $SE = 0.1947$ ,  $t = -6.3950$ ,  $p < .0001$ ). A similar but nonsignificant correlation is apparent for /k/. For the peak frequency measure, there is a significant relationship with perceptual rating for /k/ (Estimate = -22.1460,  $SE = 7.3650$ ,  $t = -3.0070$ ,  $p = .0088$ ), and a similar but nonsignificant relationship is apparent for /s/ and /ʃ/. While we have seen differences between patient groups for peak amplitude difference in /t/, neither of our acoustic measures appears to capture what the listeners were responding to when they evaluated /t/.

### Spectral Differences in Terms of Cephalometric Measures

We have observed differences according to malocclusion class. We additionally expect the severity of malocclusion to correlate with the severity of its impact on speech. Figure 10 displays the results of the random forest analyses, showing the importance of each of the

**Figure 8.** Main spectral peak frequency (Hz) for each consonant across participant groups. AOB = anterior open bite.



**Table 4.** Results of linear mixed-effects regression involving participant group differences in peak frequency (in log Hz). Reference levels are Class I, closed bite, female, /t/, and following /a/.

Predictors	Estimates	SE	Statistic	p
Intercept	8.64	0.03	268.72	< .001
Class II	-0.06	0.04	-1.30	.193
Class III	0.14	0.03	4.14	< .001
AOB	0.19	0.08	2.40	.016
Gender [M]	-0.12	0.04	-2.71	.007
Consonant /s/	0.38	0.03	14.71	< .001
Consonant /j/	-0.43	0.03	-16.61	< .001
Consonant /tj/	-0.35	0.03	-13.53	< .001
Following [æ]	0.06	0.02	3.09	.002
Following [i]	0.08	0.02	4.04	< .001
Following [u]	-0.00	0.02	-0.25	.805
Class II × AOB	-0.33	0.10	-3.23	.001
Class III × AOB	-0.19	0.09	-2.18	.029
Class II × gender [M]	-0.05	0.07	-0.70	.485
Class III × gender [M]	-0.14	0.05	-2.50	.013
AOB × gender [M]	-0.33	0.15	-2.23	.026
Class II × consonant /s/	-0.04	0.03	-1.23	.219
Class III × consonant /s/	-0.16	0.02	-6.75	< .001
Class II × consonant /j/	0.06	0.03	2.04	.041

(table continues)

Table 4. (Continued).

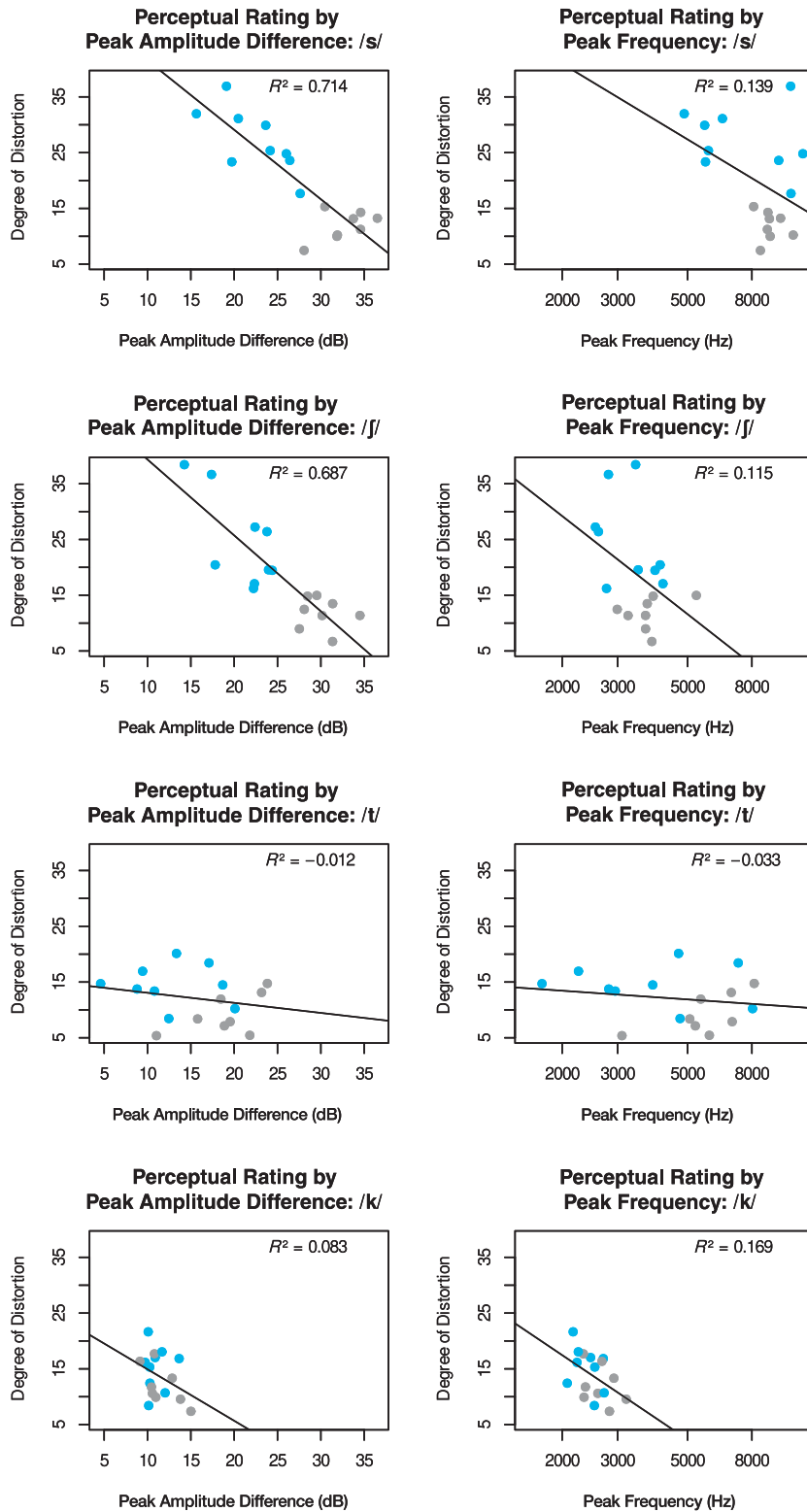
Predictors	Estimates	SE	Statistic	p
Class III × consonant /j/	-0.15	0.02	-6.47	< .001
Class II × consonant /tj/	0.06	0.03	1.92	.054
Class III × consonant /tj/	-0.11	0.02	-4.91	< .001
AOB × consonant /s/	-0.23	0.05	-4.24	< .001
AOB × consonant /j/	-0.27	0.05	-5.03	< .001
AOB × consonant /tj/	-0.21	0.05	-3.91	< .001
Gender [M] × consonant /s/	-0.18	0.03	-5.98	< .001
Gender [M] × consonant /j/	-0.12	0.03	-4.23	< .001
Gender [M] × consonant /tj/	-0.08	0.03	-2.55	.011
(Class II × AOB) × gender [M]	0.47	0.18	2.62	.009
(Class III × AOB) × gender [M]	0.27	0.16	1.71	.087
(Class II × AOB) × consonant /s/	0.35	0.07	5.00	< .001
(Class III × AOB) × consonant /s/	0.19	0.06	3.23	.001
(Class II × AOB) × consonant /j/	0.31	0.07	4.48	< .001
(Class III × AOB) × consonant /j/	0.28	0.06	4.62	< .001
(Class II × AOB) × consonant /tj/	0.23	0.07	3.27	.001
(Class III × AOB) × consonant /tj/	0.14	0.06	2.27	.023
(Class II × gender [M]) × consonant /s/	0.12	0.05	2.69	.007
(Class III × gender [M]) × consonant /s/	0.16	0.04	4.36	< .001
(Class II × gender [M]) × consonant /j/	0.04	0.05	0.99	.320
(Class III × gender [M]) × consonant /j/	0.16	0.04	4.32	< .001
(Class II × gender [M]) × consonant /tj/	0.12	0.05	2.59	.010
(Class III × gender [M]) × consonant /tj/	0.13	0.04	3.63	< .001
(AOB × gender [M]) × consonant /s/	0.25	0.10	2.46	.014
(AOB × gender [M]) × consonant /j/	0.56	0.10	5.53	< .001
(AOB × gender [M]) × consonant /tj/	0.41	0.10	3.97	< .001
(Class II × AOB × gender [M]) × consonant /s/	-0.32	0.12	-2.63	.009
(Class III × AOB × gender [M]) × consonant /s/	-0.25	0.11	-2.29	.022
(Class II × AOB × gender [M]) × consonant /j/	-0.60	0.12	-4.95	< .001
(Class III × AOB × gender [M]) × consonant /j/	-0.49	0.11	-4.43	< .001
(Class II × AOB × gender [M]) × consonant /tj/	-0.45	0.12	-3.68	< .001
(Class III × AOB × gender [M]) × consonant /tj/	-0.26	0.11	-2.32	.020
Random effects				
$\sigma^2$			0.07	
$\tau_{00}$ speaker			0.02	
$\tau_{00}$ word			0.00	
ICC			.24	
N <sub>word</sub>			16	
N <sub>speaker</sub>			284	
Observations			12,994	
Marginal R <sup>2</sup> /conditional R <sup>2</sup>			.555/.661	

Note. Values in bold are significant by  $p < .05$  convention. SE = standard error; AOB = anterior open bite; ICC = intraclass correlation coefficient.

cephalometric measures as well as gender and vowel context in accounting for each of the two acoustic measures, for each of the five consonants. The top panel shows peak frequency (which was log transformed) and the bottom panel shows peak amplitude difference. The x-axes show the percent increase in mean standard error (%IncMSE) of the predictions when the values of a given measure are randomly scrambled.

We see that cephalometric measures as a group are more important for accounting for spectral properties of /s t j tj/ than for /k/. On the other hand, vowel context is extremely important for accounting for the peak frequency of /k/, which is known to be closely related to F2 frequency (Liberman et al., 1952). It is also important for accounting for the peak amplitude of /k/. Gender is most important for accounting for peak frequency of /s/ and /j/.

**Figure 9.** Comparison of perceptual measures of speech distortion versus peak amplitude difference and peak frequency for four consonants. Blue = Class III participants; gray = Class I control participants. The distortion scale ranges from 0 (*completely clear*) to 100 (*severely distorted*).





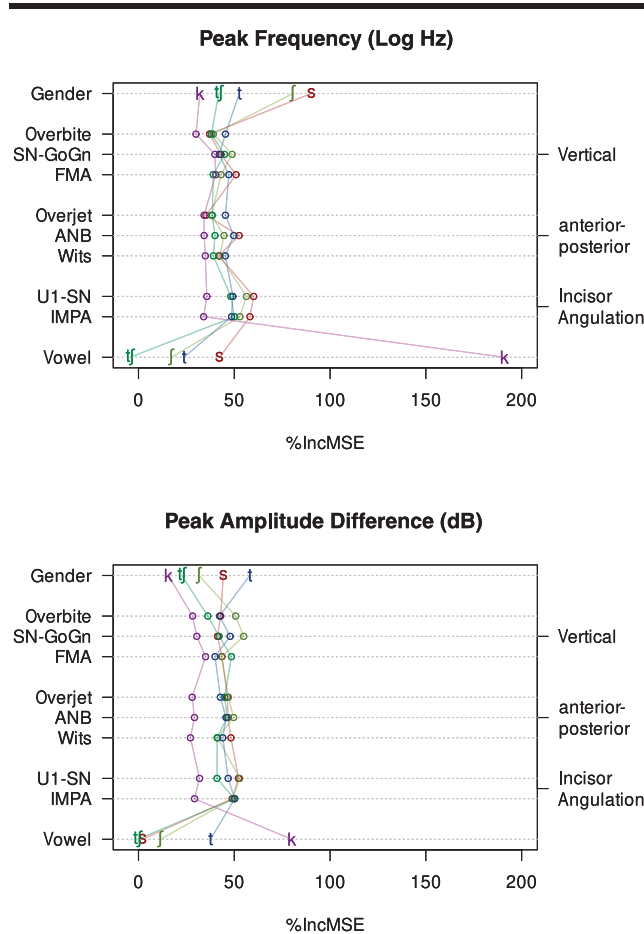
**Table 5.** Results of linear regressions examining relationships between perceptual and acoustic measures.

Acoustic measure	Consonant	Estimate	SE	t	p
Peak amplitude difference (dB)	/s/	<b>-1.2452</b>	<b>0.1947</b>	<b>-6.3950</b>	<b>&lt; .0001</b>
	/ʃ/	<b>-1.3612</b>	<b>0.2266</b>	<b>-6.0060</b>	<b>&lt; .0001</b>
	/t/	-0.1891	0.2776	-0.6810	.5062
Peak frequency (log Hz)	/k/	-0.9964	0.5116	-1.9480	.0704
	/s/	-14.8090	7.8310	-1.8910	.0781
	/ʃ/	-19.2100	10.9500	-1.7540	.0999
	/t/	-0.9756	3.5559	-0.2740	.7880
	/k/	<b>-22.1460</b>	<b>7.3650</b>	<b>-3.0070</b>	<b>.0088</b>

Note. Values in bold are significant by  $p < .05$ . SE = standard error.

For amplitude difference, gender is generally less important than the cephalometric measures. Overbite and SN-GoGn stand out as being particularly important for the peak amplitude difference of /ʃ/, and IMPA stands out as

**Figure 10.** Random forest analysis of peak amplitude difference (dB, top) and peak frequency (log dB, bottom) for five consonants. SN-GoGn = Sella-Nasion-Gonion-Gnathion; FMA = Frankfort-mandibular plane angle; ANB = A-point-Nasion-B-point; Wits = Wits appraisal; U1-SN = Upper-1 (central incisor) to Sella-Nasion Angle; IMPA = mandibular incisor to mandibular plane angle; IncMSE = increase of mean standard error.



being particularly important for the peak amplitude difference of /tʃ/ and the peak frequency of /s/.

The results of the regressions including cephalometric measures are summarized in Tables 6–7. We performed separate regressions for Class II, and Class III for all AP measures, since Class II and Class III fall on opposite sides of Class I with respect to these measures. Two of these regressions show significant effects. Overbite is positively correlated with peak amplitude (Estimate = 0.2418, SE = 0.0566,  $df = 280.1$ ,  $t = 4.2742$ ,  $p < .0001$ ), meaning that increased AOB is associated with decreased sibilance. Overjet is positively correlated with peak amplitude among Class III patients (Estimate = 0.2541, SE = 0.0669,  $df = 148.9$ ,  $t = 3.7976$ ,  $p = .0002$ ), meaning that for that subset, greater distance between upper and lower incisors is associated with decreased sibilance. SN-GoGn, FMA, and Wits (Class III) also show numerical differences in the expected directions that do not meet the significance threshold of .0021 (due to multiple comparisons).

For visualizing the effects of vertical and AP differences in patient groups relative to controls, we need to focus on the overbite and overjet measures that are present for all groups of participants. We have already seen that Class II and Class III patients, which differ from controls in opposite ways with respect to overjet and other AP measures, differ from controls in the same direction with respect to spectral peak amplitude difference. It is clear that there is not a simple linear relationship between overjet and peak amplitude difference. Figure 11 shows separate regression lines for each occlusal class.

## Discussion

This study provides quantitative assessments on a large DFD sample representing all major malocclusion classifications, stratified by AP and vertical cephalometric measures. The multitaper analysis shows differences across DFD groups that have not previously been observed.

**Table 6.** Results of linear mixed-effects regressions examining direction of effects on peak amplitude difference.

Model of peak frequency	Hypothesis	Est	SE	df	t	p
<b>Overbite</b>	<b>+</b>	<b>0.2418</b>	<b>0.0566</b>	<b>280.1</b>	<b>4.2742</b>	<b>&lt; .0001</b>
SN-GoGn	–	–0.0592	0.0242	217.1	–2.4437	.0153
FMA	–	–0.0507	0.0244	217.3	–2.0774	.0389
Overjet (Class I)	N/A	–0.3394	0.2500	65.5	–1.3575	.1793
Overjet (Class II)	–	–0.0254	0.1249	59.9	–0.2033	.8396
<b>Overjet (Class III)</b>	<b>+</b>	<b>0.2541</b>	<b>0.0669</b>	<b>148.9</b>	<b>3.7976</b>	<b>.0002</b>
ANB (Class II)	–	–0.2342	0.1220	57.9	–1.9204	.0597
ANB (Class III)	+	–0.0326	0.0613	147.7	–0.5328	.5950
Wits (Class II)	–	0.0377	0.0796	57.9	0.4732	.6378
Wits (Class III)	+	0.1279	0.0502	147.9	2.5474	.0119
U1-SN	N/A	–0.0046	0.0186	214.4	–0.2446	.8070
IMPA	N/A	–0.0237	0.0194	216.4	–1.2214	.2233

Note. Values in bold are significant by  $p < .0021$  (due to multiple comparisons). Est = estimate; SE = standard error; df = degrees of freedom; SN-GoGn = Sella-Nasion-Gonion-Gnathion; FMA = Frankfort-mandibular plane angle; N/A = not applicable; ANB = A-point-Nasion-B-point; Wits = Wits appraisal; U1-SN = Upper-1 (central incisor) to Sella-Nasion Angle; IMPA = mandibular incisor to mandibular plane angle.

First, the results of this study show that the control group produces /s tʃ tʃ/ with a higher spectral amplitude difference than any other patient group. As discussed by Koenig et al. (2013) this measurement is designed to measure the “sibilance” of fricatives and reflect variations in the noise source at different frequency regions.

With controls 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 (see Bode et al., 2023), 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 turbulence noise similar to that of the controls. The Class III patients exhibiting lower amplitude can be explained by those patients commonly having prognathic mandibles in which their tongues naturally rest more anteriorly relative to the maxilla and maxillary incisors. Adaptation for tongue position in Class III patients is difficult, potentially leading to a more anterior constriction location.

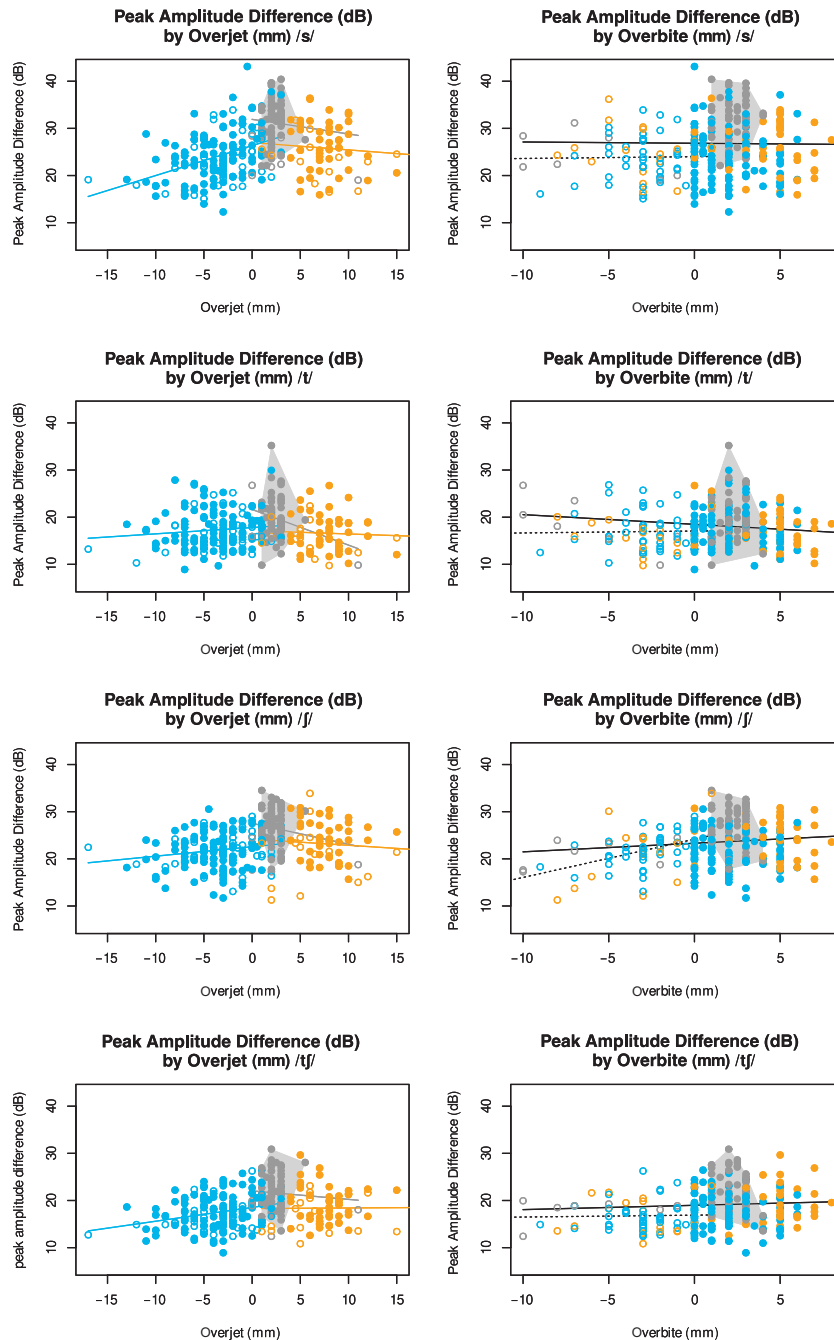
Second, the control group does not produce postalveolar consonants /ʃ tʃ/ with a different peak frequency from other patient groups, but does produce /s/ with a higher peak frequency than Class I AOB and Class II

**Table 7.** Results of linear mixed-effects regressions examining direction of effects on peak frequency.

Model of peak frequency	Hypothesis	Est	SE	df	t	p
Overbite	+	0.0014	0.0026	280.5	0.5275	.5983
SN-GoGn	–	–0.0012	0.0014	217.6	–0.8590	.3913
FMA	–	–0.0006	0.0014	217.7	–0.4073	.6842
Overjet (Class I)	N/A	–0.0204	0.0119	65.4	–1.7184	.0905
Overjet (Class II)	+	0.0055	0.0064	60.1	0.8628	.3917
Overjet (Class III)	–	–0.0004	0.0040	149.4	–0.0951	.9243
ANB (Class II)	–	–0.0006	0.0065	58.1	–0.0923	.9268
ANB (Class III)	+	0.0006	0.0034	148.1	0.1756	.8609
Wits (Class II)	–	0.0012	0.0042	58.1	0.2849	.7767
Wits (Class III)	+	0.0017	0.0029	148.3	0.5889	.5569
U1-SN	N/A	0.0003	0.0010	214.9	0.2954	.7680
IMPA	N/A	–0.0010	0.0011	216.9	–0.8855	.3769

Note. Est = estimate; SE = standard error; df = degrees of freedom; SN-GoGn = Sella-Nasion-Gonion-Gnathion; FMA = Frankfort-mandibular plane angle; N/A = not applicable; ANB = A-point-Nasion-B-point; Wits = Wits appraisal; U1-SN = Upper-1 (central incisor) to Sella-Nasion Angle; IMPA = mandibular incisor to mandibular plane angle.

**Figure 11.** Overjet is positively correlated with amplitude difference among the Class III group; overbite is positively correlated with peak frequency: /s t f tʃ/. Blue = Class III; orange = Class II; gray = Class I, unfilled = AOB. The gray regions are convex hulls around Class I closed bite. The overjet figures show three regression lines for Classes I, II, and III. The overbite figures show two regression lines for AOB (dashed) and closed bite (solid). AOB = anterior open bite.



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. Note that the compensatory behavior that would raise peak frequency is less lip rounding. The difference in the results for alveolar versus nonalveolar sounds may show that the peak frequencies of sounds with more anterior constriction locations are more affected by malocclusions than are sounds with

more posterior constrictions. When lip rounding is not a feature of the consonant, as in /t/ and /s/, all patient groups have somewhat lower peak frequencies than controls at least for /t/. Class I AOB and Class II have lower peak frequencies for /s/ as well, and this does not have an obvious explanation in front cavity characteristics, since both groups have the potential for lip incompetence. We note that Class I AOB is our smallest patient group so our findings here are least generalizable. Finally, the control group produces all consonants (except for /k/) with a higher amplitude difference than all other patient groups. These results show that producing a prominent spectral peak is difficult as overjet deviates from a normal jaw configuration (i.e., as a patient becomes Class II or Class III). We have seen that perceptual judgments of consonant distortion are most closely related to peak amplitude difference for /s/ and /ʃ/, most closely related to peak frequency for /k/, and not closely related to either acoustic measure of /t/.

Previous studies (Keyser et al., 2022; Lathrop-Marshall et al., 2022) found higher center of gravity and spectral variance 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 DFT-based analyses applied preemphasis and calculated spectral moments over a wide frequency range (0–17.64 kHz), meaning that a relatively flat spectrum can yield a higher center of gravity than one with a prominent spectral peak under 10 kHz. Using multitaper spectra to identify a spectral peak associated with a vocal tract 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 constriction at the noise source that cannot be easily compensated for with the lips.

The fact that tooth-based measures such as overbite and overjet are able to account for acoustic measures better than skeletal measures could be accounted for by the fact that the patient population is preparing for jaw surgery. A skeletal discrepancy can be quite severe but the teeth look more moderate because they are angled to accommodate the underlying skeletal position. This is called compensation. Before jaw surgery, patients undergo decompensation, in which an orthodontist moves the teeth within the jaws to prepare for how the jaws will be positioned after surgery. This has the effect of exaggerating overbite and overjet measures, because the teeth had previously been compensated for the jaw mismatch. Our pre-op sample includes a mixture of patients who have and have not undergone decompensation. As such, there is a skeletal–dental mismatch and the dental measures appear

to best account for the speech measures. This also suggests that anterior tooth positioning is more directly impactful for speech than the skeletal position itself. All of these conclusions are drawn from laboratory speech, and it would be interesting to know whether they are also observed in more natural speech tasks.

## Conclusions

It is well established that patients with DFD face challenges with the production of consonants. The goal of this study was to determine the acoustic aspect of consonant production that distinguishes among various groups of DFD patients. The analysis used two targeted acoustic measures: the main spectral peak amplitude as a measure of sibilance and the frequency of this peak as an index of the resonating cavity anterior to the exit from the constriction formed by the tongue. The differences between DFD patients and controls involve the noise source, more than the front cavity resonance. There may be more degrees of freedom available to modify the resonant frequency (such as by shaping the lips differently) than are available to increase the strength of the sibilance. Although Class II, Class III, and AOB patients differ from controls in different ways, they all produce lower-amplitude noise in /t s f tʃ/, suggesting that noise produced by the controls is near the limit of the practical range for consonant sound production, and the acoustic targets for American English consonants are well suited to typical dentofacial relationships.

## Data Availability Statement

Data are available on request from the authors.

## Acknowledgments

This research was supported by the Oral and Maxillofacial Surgery Foundation Research Support Grant (to L.A.J.) and the American Association of Orthodontists Foundation Research Aid Award. Finally, this research was funded by the National Center for Advancing Translational Sciences, National Institutes of Health (NIH), through Grant Award Number UL1TR002489 to Laura Anne Jacox, the National Institutes of Dental and Craniofacial Research, NIH through a K08 award (to L.A.J.), with a Grant Award Number 1K08DE030235-01A1, and by the National Science Foundation Social, Behavioral and Economic Sciences through grant SMA-1730479 to Jeff Mielke. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. We would like to thank all

of our participants for their time and engagement in this study. We appreciate the support of the University of North Carolina (UNC) Oral and Maxillofacial Surgery group for their input and collaboration. We value the support and guidance of S.T. Phillips in the UNC Clinical Research Unit (Go Health). We are grateful to Kendall Garland, Allyse Longrie, and Yushan Xie for their role in data analysis, Robin Dodsworth for statistics guidance, and Erika Rezende Silva for help with figures.

## References

- Amr-Rey, O., Sánchez-Delgado, P., Salvador-Palmer, R., Cibrián, R., & Paredes-Gallardo, V. (2022). Association between malocclusion and articulation of phonemes in early childhood. *The Angle Orthodontist*, 92(4), 505–511. <https://doi.org/10.2319/043021-342.1>
- Assaf, D. D. C., Knorst, J. K., Busanello-Stella, A. R., Ferrazzo, V. A., Berwig, L. C., Ardenghi, T. M., & Marquezan, M. (2021). Association between malocclusion, tongue position and speech distortion in mixed-dentition schoolchildren: An epidemiological study. *Journal of Applied Oral Science*, 29. <https://doi.org/10.1590/1678-7757-2020-1005>
- Black, L. I., Vahratian, A., & Hoffman, H. J. (2015). Communication disorders and use of intervention services among children aged 3–17 years: United States, 2012. *NCHS Data Brief*, 205, 1–8. <https://www.cdc.gov/nchs/data/databriefs/db205.pdf>
- Bode, C., Ghaltakhchyan, N., Rezende Silva, E., Turvey, T., Blakey, G., White, R., Mielke, J., Zajac, D., & Jacox, L. (2023). Impacts of development, dentofacial disharmony, and its surgical correction on speech: A narrative review for dental professionals. *Applied Sciences*, 13(9), Article 5496. <https://doi.org/10.3390/app13095496>
- Bredin, H., Yin, R., Coria, J. M., Gelly, G., Korshunov, P., Lavechin, M., Fustes, D., Titeux, H., Bouaziz, W., & Gill, M. P. (2020). Pyannote. audio: Neural building blocks for speaker diarization. In *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7124–7128). IEEE.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cronenberg, J., Gubian, M., Harrington, J., & Ruch, H. (2020). A dynamic model of the change from pre- to post-aspiration in Andalusian Spanish. *Journal of Phonetics*, 83, Article 101016. <https://doi.org/10.1016/j.wocn.2020.101016>
- Farronato, G., Giannini, L., Riva, R., Galbiati, G., & Maspero, C. (2012). Correlations between malocclusions and dyslalias: Neuromuscular evaluation in skeletal Class II and Class III patients. *European Journal of Paediatric Dentistry*, 13(1), 13–18. <https://doi.org/10.1016/j.pio.2012.04.003>
- Forrest, K., Weismer, G., Hodge, M., Dinnsen, D. A., & Elbert, M. (1990). Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clinical Linguistics & Phonetics*, 4(4), 327–340. <https://doi.org/10.3109/026992090008985495>
- Gonzalez, S., & Brookes, M. (2014). PEFAC - A pitch estimation algorithm robust to high levels of noise. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(2), 518–530. <https://doi.org/10.1109/TASLP.2013.2295918>
- Grudzią-Sękowska, J., Olczak-Kowalczyk, D., & Zadurska, M. (2018). Correlation between functional disorders of the masticatory system and speech sound disorders in children aged 7–10 years. *Dental and Medical Problems*, 55(2), 161–165. <https://doi.org/10.17219/dmp/86006>
- Jacox, L. A., Turvey, T. A., Mielke, J., Zajac, D. A., & Blakey, G. H. (2022). Impacts of jaw disproportions on speech of dentofacial disharmony patients. *Journal of Oral and Maxillofacial Surgery*, 80(9), Article S26. <https://doi.org/10.1016/j.joms.2022.07.039>
- Jesus, L. M., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, 30(3), 437–464. <https://doi.org/10.1006/jpho.2002.0169>
- Jhingree, S., Xie, Y., Bocklage, C., Giduz, N., Moss, K., Zajac, D., & Jacox, L. A. (2022). Validating spectral moment analysis as a quantitative measure of speech distortions in speakers with Class III malocclusions. *Perspectives of the ASHA Special Interest Groups*, 7(3), 728–740. [https://doi.org/10.1044/2022\\_PERSP-21-00315](https://doi.org/10.1044/2022_PERSP-21-00315)
- Keyser, M. M. B., Lathrop, H., Jhingree, S., Giduz, N., Bocklage, C., Couldwell, S., Oliver, S., Moss, K., Frazier-Bowers, S., Phillips, C., Turvey, T., Blakey, G., White, R., Zajac, D., Mielke, J., & Jacox, L. A. (2022). Impacts of skeletal anterior open bite malocclusion on speech. *FACE*, 3(2), 339–349. <https://doi.org/10.1177/27325016221082229>
- Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research*, 56(4), 1175–1189. [https://doi.org/10.1044/1092-4388\(2012\)12-0038](https://doi.org/10.1044/1092-4388(2012)12-0038)
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lathrop-Marshall, H., Keyser, M. M. B., Jhingree, S., Giduz, N., Bocklage, C., Couldwell, S., Edwards, H., Glesener, T., Moss, K., Frazier-Bowers, S., Phillips, C., Turvey, T., Blakey, G., White, R., Mielke, J., Zajac, D., & Jacox, L. A. (2022). Orthognathic speech pathology: Impacts of class III malocclusion on speech. *European Journal of Orthodontics*, 44(3), 340–351. <https://doi.org/10.1093/ejo/cjab067>
- Lieberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology*, 65(4), 497–516. <https://doi.org/10.2307/1418032>
- Lüdtke, D. (2023). *sjPlot: Data visualization for statistics in social science*. R package version 2.8.15. <https://CRAN.R-project.org/package=sjPlot>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Interspeech* (Vol. 2017, pp. 498–502).
- Morris, M. A., Meier, S. K., Griffin, J. M., Branda, M. E., & Phelan, S. M. (2016). Prevalence and etiologies of adult communication disabilities in the United States: Results from the 2012 National Health Interview Survey. *Disability and Health Journal*, 9(1), 140–144. <https://doi.org/10.1016/j.dhjo.2015.07.004>
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination  $R^2$  and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134), Article 20170213. <https://doi.org/10.1098/rsif.2017.0213>
- Natour, Y. S., & Saleem, A. F. (2009). The performance of the Time-Frequency Analysis Software (TF32) in the acoustic analysis

- of the synthesized pathological voice. *Journal of Voice*, 23(4), 414–424. <https://doi.org/10.1016/j.jvoice.2007.11.002>
- Ocampo-Parra, A., Escobar-Toro, B., Sierra-Alzate, V., Rueda, Z. V., & Lema, M. C.** (2015). Prevalence of dyslalias in 8 to 16 year-old students with anterior open bite in the municipality of Envigado, Colombia. *BMC Oral Health*, 15(1), Article 77. <https://doi.org/10.1186/s12903-015-0063-1>
- Oliver, S., Keyser, M. M. B., Jhingree, S., Bocklage, C., Lathrop, H., Giduz, N., Moss, K., Blakey, G., White, R., Turvey, T., Mielke, J., Zajac, D., & Jacox, L. A.** (2023). Impacts of anterior-posterior jaw disproportions on speech of dentofacial disharmony patients. *European Journal of Orthodontics*, 45(1), 1–10. <https://doi.org/10.1093/ejo/cjac057>
- Proffit, W. R., White, R. P., & Reinhardt, R. W.** (1991). *Surgical-orthodontic treatment* (1st ed.). Mosby.
- R Core Team.** (2000). *R language definition*. R Foundation for Statistical Computing, 3(1).
- Reidy, P. F.** (2013). *spectRum R package*. <https://github.com/patrickreidy/spectRum>
- Reidy, P. F.** (2015). A comparison of spectral estimation methods for the analysis of sibilant fricatives. *The Journal of the Acoustical Society of America*, 137(4), EL248–EL254. <https://doi.org/10.1121/1.4915064>
- Shadle, C. H.** (2006). Acoustic phonetics. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., Vol. 9, pp. 442–460). Elsevier.
- Shadle, C. H.** (2023). Alternatives to moments for characterizing fricatives: Reconsidering Forrest et al. (1988). *The Journal of the Acoustical Society of America*, 153(2), 1412–1426. <https://doi.org/10.1121/10.0017231>
- Vosk Speech Recognition Toolkit.** (2023). Retrieved April 30, 2023, from <https://alphacephei.com/vosk/>
- Winkelmann, R., & Raess, G.** (2014). Introducing a web application for labeling, visualizing speech and correcting derived speech signals. In *LREC* (pp. 4129–4133).