**Assessing articulatory perturbations during a sung vowel-matching exercise using articulography**

Author 1, Author 2

Author affiliations

Corresponding author contact details

**ABSTRACT**

Vowel matching is considered a beneficial acoustic strategy for ‘choral blend’, but little is known about the articulatory changes that occur in practice. This paper explores the use of electromagnetic articulography (EMA) to examine the articulatory modifications used by singers when asked to blend with a unison singing voice. Five choir singers were recorded sustaining vowels with EMA electrodes placed on the tongue and the lips. During this task, subjects were asked to blend to a pre-recorded stimulus (same pitch and vowel). In a control condition no stimulus was played. Vowels /ɑ, i, u/ at pitches D4, F#4 and A4 were presented in each condition in a randomised order. Fundamental frequency (fo), intensity and the first three formant frequencies were estimated from the acoustic signal and analysed alongside the synchronous EMA data. Baseline and matching conditions were compared for each recording. All subjects significantly shifted their fo in one matching condition suggesting that tuning might be their primary blending strategy. Significant differences in multiple sensor movements were only observed for the professional subject. Analysis of the trajectories of these sensor movements suggests a delay in response similar to expectation from tuning literature. EMA is shown to be a promising tool to investigate articulatory movements in addition to acoustic features of singers when blending their voice to another. These protocols will be valuable for understanding relationships between articulatory and acoustic features as voices adapt to external stimuli, which is especially applicable for vocal training and choral pedagogy.

Keywords: Singing, articulography, vowel matching, blend, choir

**1. INTRODUCTION**

In the context of choral or ensemble singing, vowel production is considered an important contributing factor to choral blend: practitioners focus on ‘vowel matching’ when training and directing choirs based on the premise that ‘intonation and overall intelligibility rely on matching vowel sounds, uniformity is crucial to producing the kind of blend that results in a truly beautiful choral tone’ [1]. As described by [2] in a review of the research in this area, empirical studies also indicate that a unified vowel is desired for a good choral sound.

Studies conducted to date tend to focus on the acoustics rather than the articulatory strategies being used to achieve ‘blend’. Hunt asserts that “the problem of unity of vowel is one of intonation of formant frequencies” [3], and a number of studies have assessed formant frequency changes in solo versus singing choral singing modes alongside other factors contributing to blend, including fundamental frequency (fo), intensity and other spectral characteristics [4, 5]. Goodwin conducted a study whereby individuals sang solo and then ‘blended’ to a pre-recorded choral ensemble delivered over headphones. Results showed a stronger intensity of fundamental frequency and first formant and weaker second and third formant in a choral condition [6]. Also utilising an external choral reference for the participant to blend with, [7] observed an increase in energy in the singer’s formant region when participants performed solo, compared to an increase in the region of the fundamental when singing with the choir. [8] used synthesis to assess the impact of changing the vowel quality of a reference stimuli on fundamental frequency agreement between singers, finding that tuning precision increased when reference vowels had common partials reinforced by formants.

The literature concerned with choir acoustics tends to focus on features of the steady state of tones such as long-term average spectra or mean fo rather than the dynamic nature of the adjustments that singers make to one another. There is, however, a related body of work that considers voice fundamental frequency response to altered auditory pitch feedback. Singing the missing third in a major triad, [9] measured the delay before participants adapted their fundamental frequency when the external pitch reference was suddenly altered. Mean latencies of 227ms for highly skilled and 206ms for moderately skilled choir singers were observed. They found the reaction time to be shorter for larger shifts in the reference and observed two phases of adjustment in some singers, which they suggest represent an initial larger adjustment followed by a, perhaps conscious, fine tuning stage. They refer to a study utilising the same approach by [10] which observed two peaks in the reaction time in professional operatic singers with latencies of 113ms and 261ms respectively, but only one peak in the distribution time of the response in non-singers at 135ms.

Other studies investigating the fo responses to pitch-shift stimuli, but of speakers [11], also observed two response phases in some subjects. In three conditions subjects were instructed to either ignore, adapt to, or respond in the opposite direction to the >500ms pitch shift in the stimulus. The latency of the first response was reduced when the participants were instructed to respond to the stimuli, although their response was sometimes in the opposite pitch direction to the instruction. When told to make a voluntary response when they heard the shift, two responses were observed but the latency of the later response depended on the condition. The authors concluded that the later response is likely voluntary, whilst the first response is automatic but can be altered through instruction. This complements the previous findings of [12] who provided participants with auditory feedback of their own voice with the pitch shifted, and found two stages of response with latencies of 160ms and 300ms respectively. Although much of this work is primarily concerned with understanding the control mechanisms of phonation frequency, which are highly accurate especially in trained singers, it is also very relevant to the larger picture of understanding the processes of singing in a choir. The fine adjustments required of choir singers when blending together demand the rapid adjustment of fo to accurately tune to one another, but also concern the fine adjustment of many more parameters, including intensity, vibrato and timbre.

It is now possible to track points on the articulators (tongue, lips, mouth) in real-time as a participant sings, with the emergence of technologies including electromagnetic articulography (EMA) [13]. This method measures the location of small sensor coils, positioned on the mouth and tongue, in 3D space over time through a electromagnetic field created by induction coils near the head. A number of studies have illustrated the potential for EMA to provide insights into the behaviour of the articulators and their contributions to acoustics during singing. Four female trained singers were recorded both singing and speaking French using EMA to track the articulatory and acoustic changes [14]. [15] used EMA on German sustained vowels and running speech to illustrate that the same vowel can be produced with very different articulator placement and that similar positions can produce two different vowel identities. The links between vowel and timbre become particularly interwoven in discussions of blend in vocal ensembles, and EMA presents a promising opportunity to investigate how the articulators are involved in the blending process.

This study provides insight into the potential for new paradigms which make use of EMA to further our understanding of the alterations singers make as they blend with other singers: the articulatory movements of the lips and tongue are investigated as a singer is tasked with matching a sung note to the audio stimulus of another unison voice. As it is the first study to employ EMA for this purpose, no formal hypothesis is presented, but the following research questions are addressed (1) Does EMA provide useful data for measuring articulation changes during a vowel matching singing task? (2) Can EMA data be mapped to acoustic parameters? (3) Can a robust and feasible experimental paradigm be developed utilising EMA and acoustic recordings to better understand the processes involved when singing voices blend together?

In the following section the methods of data collection are described, for the recording of the response stimuli and the participant data, as well as the analysis methods. The results of analysis for steady state portions of the sung tones are presented for all subjects in section 3, with observations of the transitions for subject 2. Section 4 discusses the findings and their relevance to existing literature on the adjustment strategies of singers, details the limitations of the current study and suggests areas of future work before a concluding statement in section 5.

**2. METHODS**

**2.1 Data Collection**

Data was collected for both stimulus and response recordings. Stimulus recordings were made of two biologically-defined female singers, one novice and one classically trained experienced choral singer, sustaining vowels /ɑ, i, u/ at three pitches (D4, F#4, A4) for five seconds at a moderate volume (mezzo forte). These recordings were then processed to remove the first and last second, resulting in a three-second, approximately steady-state stimulus signal for each vowel-note combination.

Response recordings were made of five experienced female choral singers aged 24-37, who did not record the stimuli. All subjects sang in choirs with membership by audition at least weekly. Subject S2 regularly performs in professional choirs and S2, S3 and S5 each had two degrees in music with voice as their first study. Subject S4 had an undergraduate music degree but no formal singing training, and Subject S1 had five years singing lessons but no formal music qualifications.

Subjects were sat within a Carstens AG501 articulograph in a rectangular room with no acoustic treatment. In both recording scenarios, EMA sensors were placed in the midsagittal plane on the upper and lower lips (sensors UL and LL), and tip and back of the tongue (sensors TT and TB); in addition, a sensor was positioned on the side of the tongue (TS) and at the corner of the mouth (LC). A small lavalier microphone (DPA 4062-BM) was taped to the subject’s nose. Subjects were also fitted with electrolaryngograph (Lx) electrodes to allow parallel measurement of vocal fold activity. Open-back headphones were worn (Beyerdynamic DT990 Pro) in the response recordings only.

The response subjects sang each vowel-note combination for six seconds at moderate volume (mezzo forte). The initial three seconds provided baseline data (condition B). After approximately three seconds they heard either no external sound (condition M0), the novice singer (condition M1) or the trained singer (condition M2) and were asked to blend, with a focus on matching their vowel, to what they heard. They were instructed on the vowel and given a reference pitch from an electronic pitch pipe before each recording. Stimuli were presented at a level calibrated to match that of the target singer in the same space. Condition, note and vowel order were randomized.

Ethical approval for this study was granted from the University of York Physical Sciences Ethics Committee with written informed consent obtained from all participants before taking part.

**2.2 Data Analysis**

Synchronised audio channels from the laryngograph, nose microphone, room microphone and stimulus recording were combined into a 4-channel, 16-bit, 48kHz WAV file. The nose microphone and stimulus channels were processed with Praat [16] to extract the pitch, and the frequencies of the first three formants. Formants were calculated using standard LPC settings with 6 formants and a maximum value of 5500Hz, and extracted automatically using Praat’s formant tracking algorithm with subject- and vowel-specific reference formants, frequency and transition costs of 10 and a bandwidth cost of 1.

For each recording, approximately one second of steady-state audio (comprising an integer number of vibrato cycles where present) was identified from the middle of both the baseline and matched singing conditions. These segments were extracted from the audio along with corresponding segments from the pitch, formant, and EMA sensor traces. Average values of each parameter across each segment were calculated, using the median for the formant values to account for occasional fluctuations due to tracking error, and using the mean for all other parameters.

Absolute differences in average parameter values between the baseline and matching conditions were then calculated. In order to maximise statistical power, results across all vowels and notes are combined for each subject, and the absolute difference is used to account for the fact that the direction of parameter change may be different for different vowels or notes. Results are calculated within-subject only since the EMA data does not account for differences in vocal tract size or articulation range between subjects, sensor placement may have varied slightly between subjects, and because subjects may have used different articulatory strategies to achieve the same goal.

Absolute differences in average parameter values from the baseline in each condition were compared using a one-way ANOVA with three levels, M0, M1 and M2 (d.f. = (2,23), 9 observations per level), followed by a Bonferroni multiple comparison test. Note that due to issues during data collection, two observations are missing for Subject 1 and one observation is missing for Subject 5.

In order to further investigate the subjects’ behaviour immediately after onset of the stimulus signal, EMA data, WAV files, and associated pitch and formant traces for the entirety of each recording were read into MATLAB. Stimulus onset time was automatically determined as the point at which the amplitude of the stimulus signal exceeded 10% of its maximum variance. Additionally, for Subject 2 only, approximate start and end times of the articulatory transition period were visually determined for each sensor in the EMA data, using MATLAB’s interactive plotting tools in order to provide approximate values for transition onset time, duration, and size.

**3. RESULTS**

**3.1 Steady State**

The differences in average feature values between steady-state sections from the baseline (B) and matching (M0/M1/M2) conditions were calculated as described in the previous section and compared within subject. The absolute difference in a feature’s average value between baseline and matching conditions will be denoted by the subscript *diff*, e.g. fo,diff. Significant differences in steady state values between conditions are summarised in Table 1 below.



**Table 1**: p-values of significant differences in features observed between matching conditions M0, M1 and M2.

***3.1.1 Acoustic Features***

Significant differences in fo,diff were observed for Subjects 1 and 3 for condition M0-M2 and subjects 2, 4 and 5 for condition M0 -M1. Subject 3 was the only subject with a significant result for Intensity,diff observed only in the M0-M2 condition.

No significant differences were observed for formant frequencies between matching conditions for any subject. Further analysis of the formant tracking of the first two formants in Praat revealed a high likelihood that the LPC algorithm is locking to the first two partials, due to the high spacing of harmonics at high fos. Small changes occurring in the first two formants may be occurring that the current analysis cannot identify.

***3.1.2 Articulation Features***

Results for the EMA sensor data are provided in three dimensions: x (posterior-anterior), y (left-right) and z (inferior-superior). Some significant changes in sensor positions were observed in all subjects apart from subject 3, although only for one sensor and one condition for Subjects 1 (M1-M2, ULy) and 5 (M0-M1, TSx) and two sensors and one condition for subject 4 (M0-M2, TSx LCy). There were no significant differences in position for the tongue back sensor for any subject. Subject 1 presents significant differences only for lateral movement of the upper lip sensors without corresponding x- or z-direction movement, suggesting consistent but asymmetrical adaptation to the stimuli.

Subject 5 presents one articulatory adaptation in addition to the change in fo,diff, with adjustments in the x-direction made in response to condition M2. Both Subjects 4 and 5 present differences in the x-direction TS sensor position without accompanying movement of any other tongue sensor; which is difficult to interpret but may simply reflect idiosyncrasies in the subjects’ articulation strategies.

Subject 2 shows the most articulatory differences during the matching conditions, with significant differences observed in LLx, LLz, LCx, LCy, LCz, for M0-M1 condition and LLz and TTx sensors between M0-M2 but no significant differences observed between M1-M2.

The absolute displacement in sensor position which constitutes a significant change is inherently variable with subject and sensor. For Subject 2, for instance, the threshold of significance for p<0.05 was found to be a difference between mean steady-state sensor values of approximately 0.95mm for LLz, 0.22mm for LLx, 0.61mm for TTz, and 0.46mm for TTx. Although these thresholds appear small, these differences apply only to the current situation--mean sensor position calculated over approximately one second of steady-state material.

**3.2 Transitions**

In addition to comparing acoustic and articulatory features once the subject has adapted to the external stimulus, as above, it is of interest to investigate how these features change dynamically after exposure to the stimulus. Figures 1-3 illustrate several examples of this transition for Subject 2, for vowel /u/ and pitch F#4 in all three matching conditions. This section will focus primarily upon the transitions for Subject 2 since, as demonstrated in the previous section, this subject demonstrated the most significant articulation changes after matching. Figures 4 and 5 provide examples of transition trends from different subjects: the consistent large fluctuations in sensor movements for Subject 1 (Figure 4) irrespective of condition are typically observed for Subject 1 but unusual in other subjects; other subjects present visible articulatory shifts with a similar latency to those observed for Subject 2 (such as Subject 5 in Figure 5), however these are not consistent enough to reach significance in the steady state analysis.



**Figure 1**: (a) fundamental frequency and (b, c,d) EMA sensor values (all but fundamental frequency have been normalised by subtracting their mean value) for Subject 2, vowel /u/, pitch F#4, in condition M0.



**Figure 2**: (a) fundamental frequency and (b, c,d) EMA sensor values (all but fundamental frequency have been normalised by subtracting their mean value) for Subject 2, vowel /u/, pitch F#4, in condition M1. ‘S.O.’ denotes stimulus onset time.



**Figure 3**: (a) fundamental frequency and (b, c,d) EMA sensor values (all but fundamental frequency have been normalised by subtracting their mean value) for Subject 2, vowel /u/, pitch F#4, in condition M2. ‘S.O.’ denotes stimulus onset time.



**Figure 4**: (a) fundamental frequency and (b, c,d) EMA sensor values (all but fundamental frequency have been normalised by subtracting their mean value) for Subject 1, vowel /α/, pitch F#4, in condition M0.



**Figure 5**: (a) fundamental frequency and (b, c,d) EMA sensor values (all but fundamental frequency have been normalised by subtracting their mean value) for Subject 5, vowel /i/, pitch D4, in condition M1. ‘S.O.’ denotes stimulus onset time.

Of the 18 recordings for Subject 2 where external stimuli were presented (conditions M1 and M2), 16 showed evidence of a visible shift in articulator movement in response to the stimuli. Initial articulator transitions had a mean duration of 0.52s, started a mean of 0.61s after stimulus onset, and spanned a mean distance of 1.74mm, with the lower lip moving most overall (mean 3.11mm) and the tongue back moving the least (mean 0.82mm). In five cases, these initial transitions were followed by articulator movements in the opposite direction, a possible response to articulatory overshoot. In two cases, initial transitions were followed by further articulator movements in the same direction as the initial transition, indicating initial articulatory undershoot which is subsequently corrected. Examples of such transitions are illustrated in Figure 6, including examples of how initial transitions were annotated.



**Figure 6:** examples of (a) “normal” transition (vowel /u/, note F#4, condition M2), (b) overshoot (vowel /α/, note A4, condition M1), and (c) undershoot (vowel /u/, note D4, condition M2) for LLZ for Subject 2. Dashed lines indicate stimulus onset; dotted lines indicate where the initial transition was marked as starting and ending.

It is likely that transition features (onset time, size, and duration) interact in some way with matching condition and vowel. However, given the present sample size and the fact that the experimental protocol did not incorporate a systematic test for such relationships, no significant correlations were found, and this is proposed as an area for future research.

Whilst striking visible shifts such as those observed for the EMA data are not obvious in the trajectories for fo or formants, there is evidence of a very gradual fine tuning of fo to the stimulus for some notes (Figure 3), although no onset for these transitions can be identified.

**3.3 Summary of results**

At least one significant result was observed for each subject, suggesting some form of adaptive behavior to the stimuli. The tracking of the formant frequencies by Praat was not deemed reliable due to the high fundamentals used in this study, with evidence of the LPC algorithm locking into the partials. Therefore formant data is not considered further Four subjects produced a significant shift in at least one articulation and one condition but there are no patterns between subjects for the points of articulation found to significantly alter when matching. The absence of patterns in points of articulation between subjects is unsurprising considering the different physiology of individuals and the different strategies that might be employed for a similar acoustic outcome.

**4. DISCUSSION**

**4.1 Overall Findings**

There are very few points of articulation with significant differences between the baseline and matching condition for all subjects apart from Subject 2 (which is discussed below).The lack of significant movement of the back tongue sensor may be indicative of there being little difference in F2 although other measurement techniques would need to be employed to test this. The unreliable tracking of formants at higher fos limits any relationships that might be observable between the EMA and acoustic data, and prevents the second research question (Can EMA data be mapped to acoustic parameters?) from being fully explored within this study. However, as discussed in section 4.3, future work could address these issues.

For those articulatory points that significance was reached there are no patterns to observe between the subjects: Subject 1 alters lower lip position between matching conditions (M1-M2), Subjects 1, and 5 move a single (but different) articulatory point point in one condition only whilst Subject 4 adjusts points on both the lips and tongue but only in one direction for one condition.

 The absence of patterns across subjects in terms of which features moved significantly for which condition is unsurprising due to the differences between the recorded stimuli; they were recorded as solo performances and therefore the stimuli presented in condition M2 (trained singer), have a deep, regular vibrato and are louder than those presented in M1 which has no vibrato (novice singer). In light of this it is perhaps surprising that more significant results are not observed in the intensity analysis.

Significant results for fo,diff for all subjects might suggest that the subjects are mainly using phonatory rather than articulatory strategies, prioritising tuning, to optimise blend, in spite of the instruction to ‘focus on matching the vowel’. The absence of any significant articulatory changes for subject 3 but significance in matching M0-M2 for both fo,diff and Intensity,diff suggests that attempts to match to the M2 stimuli were acoustic-led based on matching loudness and tuning rather than articulatory alterations. In answering research question 1 (Does EMA provide useful data for measuring articulation changes during a vowel matching singing task?), the data from this study clearly points to observable and statistically significant alterations taking place between matching conditions, whilst highlighting the expected individual nature of articulatory changes during the task. As such, this study illustrates the potential for EMA to be a very useful tool in studying articulatory movements of singers during blending tasks.

The protocol itself was very time consuming, with each session taking on average 1.5 hours, with a further hour of preparation and cleaning required by the researcher (likely to only be further increased in future, in light of precautions surrounding the COVID-19 pandemic). The process of attaching the electrodes can also be uncomfortable for the participant, and there is a limitation on the time period over which usable data can be collected before the electrodes fall off. However, with careful design the experiment was found to be repeatable and feasible to run with multiple subjects. The observed differences in both acoustic and articulatory data in the subjects of the study indicate a positive response to research question 3 (Can a robust and feasible experimental paradigm be developed utilising EMA and acoustic recordings to better understand the processes involved when singing voices blend together?), with modifications and improvements to the study presented here suggested in section 4.3.

**4.2 Subject 2**

Subject 2, the most experienced singer, is an interesting case study to explore the nature of the transitions in articulation as they produce significant differences in the most articulatory features across conditions in the steady-state analysis. .

The adjustments reflected in the statistical analysis are clearly visible as shifts in the articulatory points when observed over time (see Figures 2&3) with a delay between the introduction of the stimuli and the response of the singer. The mean latency of the initial response, measured as 610ms, is much longer than the latencies linked to an automatic response of less than 300ms that have been observed in pitch-shifting studies of singers [9, 10] and non-singers [11,12]. There are examples of a second later adjustment performed by Subject 2 (Figure 6), which is a feature also commonly observed in the pitch-shifting studies and attributed to a conscious response, functioning as ‘fine tuning’ [9]. In this study the subject is ostensibly singing the same note as the stimulus (they were given the same reference pitch) rather than adjusting to a large obvious pitch-shift as in the pitch-shift studies, and so only small adjustments to blend to the reference are expected, and in the case of fo only very small changes to refine pitch matching for singing in unison. [9] also observed longer latencies with smaller pitch shifts, although the smallest shift tested was quarter of a tone. The example latency of a second response provided by [9] is 1290ms after a 220ms initial response which is well within the initial latency observed in the current study.

Interestingly the onset of transitions in the EMA data are usually easily discernible (see Figures 2 &3), however the adjustments made to fo are generally gradual and no onset of the response can be determined. This is likely connected to the very small adjustments that the singer is making to tune to the stimulus, but also aligns with the findings of [9] who could not identify the onset of change in fo in their pitch-matching study for some singers because it was too gradual, especially for the highly skilled singers.

The similar patterns emerging in the adjustments of the articulators, including clear moments of transition, double responses and overshooting (as illustrated in Figure 6), suggests similar behaviour to the pitch-shift studies with the longer latencies suggesting that they are part of a conscious fine-tuning process. The absence of a rapid adjustment is likely to be related to the articulator muscles having different response times and control mechanisms compared to the muscles controlling phonation frequency.

**4.3 Limitations**

The greatest limitation of this study is the sample size. In addition to more subjects, it would be useful to collect repetitions of the same tasks from the same subjects. The combination of vowels and pitches included in the statistical analysis may dilute interesting results, but owing to few subjects it was not possible to interrogate the data at the vowel and pitch level with any confidence. Due to the time-consuming protocol of attaching sensors to the subjects, and the limited time that sensors remain attached, this was not possible using the current protocol (which required 1.5 hours per subject). It is likely that the differing backgrounds of the singers impacted the results - the professional subject (Subject 2) presented the most significant articulatory differences during matching, but without richer data it is not possible to theorise that this may be due to training.

A constant high-pitched hum could be heard over the headphones once in the Articulograph due to the magnetic field. However, the use of headphones was necessary to isolate the stimulus recording from that of the participants. Subjects were given time to listen to the background noise and assess its impact on their performing the task. Despite claims that it was not a problem this cannot be ignored as a confounding factor. The room in which the experiment took place is not acoustically treated and is quite ‘lively’. To mitigate against this to some extent, the stimuli were also recorded in the same room and the microphone was mounted as close as possible to the performers’ mouths (attached to the nose) to minimise the effect of the room on the recordings.

Formant extraction was performed automatically in Praat, using subject-specific settings to obtain the best performance possible, and it is very likely that the first and second formants are locking into the partials. The limits any assessment of relationships between articulatory and formant movements. Due to the sparsity of harmonics at high fundamental frequencies, acoustic analysis alone is unlikely to produce robust representation of vocal tract resonances in the output spectrum illustrative of formant frequencies. This could be resolved in future studies through the use of male voices at much lower fundamentals, or with an additional protocol added to the method to include the direct measurement of vocal tract resonance (e.g. [17]). However this measure could not be taken simultaneously, and would add even more complexities to an already time consuming experiment.

For this pilot study, sensor positions were chosen to explore a range of articulatory areas that might be of interest for future study. In future, sensor placement will be focused on specific areas of articulation; for instance, using four sensors on the midsagittal line of the tongue, working backwards from tip as a reference point. This would also improve the accuracy of the placement of the sensors on the correct point on the tongue, which was quite variable across subjects in the current study due to there being no absolute point of reference.

**4.4 Future Work**

This study has shown there is much potential for EMA to be a very useful tool in understanding the movements of the articulators when adapting to audio stimuli of other singers presented over headphones . Future work should focus on collecting reliable data, including reference points for electrode placement and with repetitions, from subjects with consistent backgrounds (in terms of training / experience) and methods to accurately measure the formant frequencies of higher voices as discussed in the limitations section. More data to allow comparison of articulation between subjects, accounting for individual vocal tract shape and size, is also considered an important area for future development and techniques for exploring this are currently under investigation by the authors.

In order to understand the implications of the articulatory and acoustic adjustments on the blending processes a follow-up study presenting the combined stimuli and responses to expert listeners would contextualise the findings in terms of perceptions of blend and help to translate the emerging findings into useful advice for singers.

**5. CONCLUSIONS**

This study investigated changes in acoustic features and points of articulation on the tongue and lips of participants as they sang sustained notes and responded to a unison auditory stimulus they were asked to ‘blend’ to. The results demonstrate that EMA is a viable and promising tool in assessing articulatory modifications of singers in blending tasks presented over headphones.. However, numerous technical details need to be considered, especially the delivery of the vowel matching stimuli, choice of sensor positions, and accurate assessment of formant frequencies for higher voices. This study has laid the foundations for further work using EMA to assess blending strategies between singers, especially highlighting the importance of considering dynamic shifts as well as steady state changes. Further work employing the methods tested here will contribute to understanding of voice acoustics, particularly relationships between the voice articulators and acoustics which might inform articulatory approaches to synthesis, as well as providing a new perspective of the skills involved in ensemble singing that could go on to inform training and practice.

**6. ACKNOWLEDGEMENTS**

The authors would like to thank the subjects who took part in the study, and Philip Harrison for his support with the praat scripts used for analysis.

**7. REFERENCES**

[1] J. Tschiggfrie, Teaching Tricky Diphthongs to Your Choir. 2015. Accessed on 06.29.2019. Available: https://www.smartmusic.com/blog/teaching-trickydiphthongs- to-your-choir/.

[2] S. Ternström, “Choir acoustics: an overview of scientific research published to date,” Int J. Res. Choral Singing, vol.1, pp.3-12, 2003.

[3] W.A. Hunt, ‘Spectrographic analysis of the acoustical properties of selected vowels in choral

sound’, EdD thesis, North Texas State Univ. 1970.

[4] S. Ternström, and J, Sundberg, “Formant frequencies of choir singers,” J. Acoust. Soc. Am., vol. 86(2), pp.517-522, 1989

[5] S. Ternström, “Physical and acoustic factors that interact with the singer to produce the choral sound,” J. Voice, vol. 5(2), pp.128-143, 1991

[6] A. W. Goodwin, “An acoustical study of individual voices in choral blend,” J. Res. Music Educ., vol. 28(2), pp.119-128, 1980.

[7] T. D. Rossing, J. Sundberg, and S. Ternström, “Acoustic comparison of voice use in solo and choir singing,” J. Acoust. Soc. Am., vol. 79(6), pp.1975- 1981, 1986.

[8] S. Ternström, and J. Sundberg, “Intonation precision of choir singers,” J. Acoust. Soc. Am., vol. 84(1), pp.59-69, 1988.

[9] A. Grell et al., "Rapid pitch correction in choir singers." J. Acoust. Soc. Am., vol 126(1), pp. 407-413, 2009.

[10] C. Kestler, "Experimentelle Studie zur schnellen Tonhöhenkorrektur bei Sängern und bei Nichtsängern." (1999).

[11] T.C. Hain et al., "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex." Experimental Brain Research, vol 130 (2), pp. 133-141, 2000.

[12] T. A. Burnett et al., "Voice F0 responses to manipulations in pitch feedback." J. Acoust. Soc. Am., vol 103(6), pp. 3153-3161, 1998.

[13] P. W. Schönle et al., “Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract,” Brain Lang., vol. 31(1), pp.26-35, 1989.

[14] K. Evgrafova et al., "The study of acoustic-articulatory relations in producing singing vowels with the use Of EMA." in *Proceedings and Report-9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, MAVEBA 2015.* Firenze University Press, 2015, pp.95-98.

[15] D. Maurer et al., "Re-examination of the relation between the vocal tract and the vowel sound with electromagnetic articulography (EMA) in vocalizations." Clinical linguistics & phonetics vol. 7(2), pp.129-143, 1993.

[16] P. Boersma, “Praat: a system for doing phonetics by computer,” Glot International, vol. 5(9/10), pp.

341-345, 2001.

[17] N. Henrich N et al., “Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones.” J. Acoust. Soc. Am., vol 129(2), pp.1024-35, 2011.