

An Investigation in the Measurable Differences between Pitch Perception in the Voice and Pitch Perception of External Sound Sources

Thomas J. Edmonds, and David M. Howard, *Egham, United Kingdom*

Abstract: Objectives. Pitch perception is an important part of accurate singing. Therefore, accurate singing requires the ability to accurately assess the pitch in one's own voice. There are two objectives of this study the first was to investigate whether there is a measurable difference in perceived pitch in one's voice to the pitch one perceives from an external sound source. The second, to measure the effects of occlusion on pitch accuracy over a melodic phrase.

Study Design. We recruited 16 participants for this study. The study that was designed to investigate the perceptual difference was split into two parts. The first is a one-to-one pitch matching test where they would recreate a pitch with singing and matching external pitches. The second was singing the familiar song 'Happy Birthday' which was used to measure pitch accuracy over a melodic phrase and to measure the effects of occlusion on pitch accuracy while singing.

Methods. The one-to-one study involved singing back a series of 5 notes to a set vowel which were the same 5 notes used when matching them to a series of possible pitches on the button test. The melodic test was to sing 'Happy Birthday' 3 times, the first normally, the second wearing headphones to occlude the ear to reduce air conductive hearing and the third time with white noise to mask all hearing.

Results. The results showed a higher accuracy of pitch matching with external sounds over using their voice, and some form of occlusion (wearing headphones or headphones with white noise) showed the version with higher pitch accuracy.

Conclusions. The results of this study showed that there was improved pitch accuracy when comparing two external sounds in pitch and when singing occlusion of some form improved pitch accuracy. This could suggest a difference when recreating pitch between the voice and matching external sound sources. Furthermore, with the improvements shown from occluding the ears, it could further suggest a difference in pitch perception abilities between the voice and external sound sources. This could have implications of improving pitch accuracy in a studio environment.

Key words: Fundamental frequency—Air-conduction—Bone-conduction—Cents—Button test—Pitch matching.

INTRODUCTION

The physical act of singing requires a source of energy; which is airflow; the sound source (or oscillator); which are the vocal folds; and the sound modifiers; which are the vocal and nasal tracts.^{1,2} Being 'in tune' is a basic requirement of 'good' singing, and accurate perception of pitch by singer and listener is necessary to underpin and monitor excellent tuning when singing.

Because pitch is a psychological attribute of voice as well as a physical one, the ability to characterize vocal pitch can be done through both subjective perceptual evaluation and objective acoustic assessment.³ Successful singing requires accurate voice perceptual skills, and to be able to sing accurately a singer needs to perceive their vocal pitch accurately.⁴ Loui et al.⁵ suggest that there are amusical individuals whose

pitch production abilities do not necessarily reflect their pitch perception abilities and vice versa which could suggest that there can be a disconnect between being able to hear what is in tune and being able to sing in tune.

The hearing system is an important element in perceiving and recreating pitch accurately. The fact that the larynx is internally linked via bone conduction to the ear canal means that 'internal head transmission paths give colouration to the sound of one's voice that is quite different from the colouration that arises from external airborne paths around the head and within room and concert hall enclosures.⁶⁻⁸ The common experience that people perceive themselves as sounding different to how they expect when listening to a recording of their voice is because bone and muscle conduction filter out higher frequencies of the voice. When speaking or singing the sounds is quite directional and radiated from the nose and lips, at a meter distance at frequencies from C4 (261.6Hz) to C5(523.3Hz), 'the intensity of the sound at the ears is less by half than that directly in front of the mouth'.⁹ At these frequencies, bone conduction will be more important when hearing one's own voice because the sound travelling to the ear will be at a reduced level compared to the sound travelling from the nose and/or mouth. For higher frequency notes

Accepted for publication November 18, 2022.

From the Royal Holloway, University of London.

Address correspondence and reprint requests to Thomas J. Edmonds Royal Holloway, University of London, Egham hill, TW20 0EX E-mail: edmonds.thomas94@gmail.com

Journal of Voice, Vol. ■■■, No. ■■■, pp. ■■■-■■■
0892-1997

© 2022 The Authors. Published by Elsevier Inc. on behalf of The Voice Foundation. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.1016/j.jvoice.2022.11.026>

between 'C7 (2093Hz) and C8, (4186Hz) the region of the singer's formant, the sound at a meter in front of the mouth is twice as loud as that at one meter from either ear'.⁹ The tuning system used in reference to the musical notes is based on A440 in equal temperament. This would suggest that the sound received from AC (Air conduction) does not give the singer an accurate or full representation of the sound they are emitting due to a loss in sound intensity of their own voice.

Howard and Angus⁴ conducted a listening test in which they found that trained singers could not discriminate better than 20 cents. This could be one definition of what is classed as in tune if a change in pitch cannot be detected. Small changes in pitch less than 20 cents could be described as timbral⁴ otherwise known as tone colour or tone quality. However, Sundberg¹⁰ argues that '[i]nvestigations have revealed that [...] [p]honation frequency errors exceeding 3 cents are detected by the most skilled musical ears, and musically trained listeners hear errors of 5 cents corresponding to about 0.3% of the frequency'.¹⁰ A variance from a target note of 100 cents would be a semi-tone. In a study by Cobes, she uses the term 'uncertain singer'¹¹ for those of her subjects, 'who sang three uncertain pitches out of four'. The criterion for an 'uncertain' pitch was a sound, '...at least a half-step [50 cents] sharp or flat'.¹¹ The importance of using cents when assessing f_o is because it is a logarithmic scale of measurement between two frequencies, whereas Hz is exponential. The intervals between notes in Hz grow larger the higher the notes. This would make a comparison between a low male voice and a high female voice misleading in results when comparing pitch accuracy. In this instance, 100 is equal to a semitone in western classical music and the unit of measurement remains the same regardless on the height of the notes.¹²

Loui et al.⁵ evaluated the relationship between vocal pitch-matching skills and pitch perception skills in untrained accurate and inaccurate singers. One of a small number of similar studies to the current study that compares the relationship between pitch perception accuracy and pitch-matching accuracy.

In a study conducted by Watts et al. it was found that pitch discrimination accuracy was the cause of more than 40% of error in a pitch matching task.¹³ These results highlighted the importance of the relationship between pitch discrimination, or perception and pitch-matching. To accurately match a target pitch, the singer must simultaneously coordinate their auditory perception and their motor abilities of using their vocal mechanism to reproduce the perceived pitch.¹³

The 'sensation of the voice is a result of the transmission of frequency through the bone and muscle of the head. This transmission via bone and muscle conduction filters the higher frequencies creating a bias toward the lower frequencies.⁹ This 'darker' sound bias in an individual's voice suggests that there is an existing perceptual difference in how we perceive our voices and how we perceive external or stimulus pitch.

Von Békésy reported that these internal paths of acoustic energy transmission were responsible for about half of the self-perceived loudness of the speaking and, by implication, the singing voice.¹⁴ This suggestion indicates a balance to how we hear our voices when we sing and subsequently, pitch. However, '[p]oor pitch singers may not necessarily be deficient in pitch discrimination, but rather have inadequate control over the voice mechanism in reproducing vocally a given pitch'.¹⁵

This paper investigates measurable differences between how the pitch of a singer's voice is perceived by themselves compared to the pitch of an external sound source. The ability to assess the accuracy of one's own singing is essential to good singing. Pitch perception accuracy is a basic requirement in singing and for the purposes of this work, it is defined as a particular quality of a sound (e.g., an individual musical note) that fixes its position in the scale.¹⁶

Pitch perception can also be described as different notes played or sung at the same volume a change of pitch can be detected. This is generally described as raising or lowering pitch which gives the listener the perception of higher and lower pitches.¹⁷ The practice of ear covering is commonplace in choirs and vocal studios to help a singer 'tune in' to their own sound more. This study also measures and assesses the effects that ear canal occlusion (by wearing headphones or ear defenders) can have on pitching accuracy when singing a melody of the familiar tune 'Happy Birthday'. The literature has stated that good singing requires accurate perceptual skills and when this isn't present it can result in singing out of tune whilst being unaware. These observations have also been noted in the authors' experience as a voice teacher and singer. The purpose of this investigation is to provide evidence to enhance overall understanding of pitch perception in singing and to provide strategies to improve vocal pitching accuracy in practice.

Two themes of this literature to consider in relation to this study is that perceptual skills are important in the process of pitch matching and that vocal coordination can be a factor in poor pitch singing separate to perceptual skills.

METHODS

Pitch discrimination accuracy¹³; is a basic skill and two tests were devised firstly to assess pitch discrimination accuracy and secondly to see how the addition of ear canal occlusion of a singer to reduce air-conduction (AC) affects the pitch perception of that singer and its overall effect on tuning accuracy. In these two tests, an occlusion of the ear canal was created to reduce air-conduction and enable the measurement of any resulting change in accuracy of f_o (fundamental frequency) of the singer's sung notes and compare that occluded version to their normal singing.

Procedures for data collection

As part of the study, participants completed a questionnaire to give information on age, gender, any knowledge of hearing loss or impairment, headphones used in the study, and

to rank their sung results. Screening of hearing loss or impairment in the questionnaire was to exclude outliers. No participants in this declared any hearing loss or impairment. All singers in this study were choral singers.

At the time of the study, Covid-19 restrictions meant that conducting the study in person wasn't possible. This meant that data collection process was arranged to be carried out remotely based on clear instructions given to the participants to minimise the variation of the data sets recorded on different devices and microphones in different environments. A study which compared the use of a smartphone microphone against specialised microphone found that 'Calculation of f_o [...] [was] not significantly different when comparing the choice of microphone in this study'.³ However, as the recordings generated results relative to itself and not directly compared to another recording, it was decided that the variable factors would not be detrimental to the results. The study was conducted via Zoom¹⁸ and prior to the study taking place, a series of checks were done to test the quality of the microphone and the stability of the internet. Set up checks were done in the Zoom settings to enable original sound for musicians, high fidelity music mode and ensure suppressing background noise was on low to ensure optimal recording settings. This was done on an individual basis based on the participants equipment by singing something short that they knew already or singing a single note loudly or at a high pitch. This was to check the individual microphone didn't clip any of the sound and disrupt the recording. In most cases, there was only one microphone option but where there was an additional option it was made sure that the same was used throughout.

All the sound files were converted to WAV from MP3 and M4A that were taken from a Zoom¹⁸ recording, mobile phone or tablet or computer using Audacity before being analysed for f_o using Praat.¹⁹ The frequency analysis range was set in Praat¹⁹ to include f_o ; clipped data or formants would therefore not be excluded from this analysis range. As this study is only analysing f_o and not spectral properties of sound, data samples were not affected by the variable factors of recording of different recording spaces, different equipment and different original file types.

Oloza et al.²⁰ found that the use of smartphones to be reliable in clinical settings for measurements of acoustic voice parameters. These findings, deemed it feasible to accept recordings made on smartphones. They were then asked to place their phone or recording device between 30 to 50cm from mouth to microphone and maintain this distance for the duration of the recording. This distance is based on what Titze describes as a commonly used distance²¹ (30cm) with the additional margin if the proximity to the microphone was causing the sound to distort. Instructions for this were given in a participant information document prior to the recording.

Issues that can arise when measuring pitch accuracy can relate to data collection and the tests implemented to assess pitch accuracy. Some studies measures only single pitch

matching to a piano timbre, whereas others have participants singing a song from memory.⁵ There are many different ways in which singing can be assessed. It was important for this study to have some sort of calibration between perception and recreation of pitch as well as a more realistic representation of singing. This study was divided into two tests to assess a range of skills needed for singing in tune. Test 1 was a one-to-one pitch matching test designed to measure pitch perception and recreation using the voice and an external sound generator. This is because 'successful singing requires perceptual skills'.⁴ Five sampled piano notes are played and after each note, the subject was asked to sing back the note to an 'i' vowel (see Table 1) after each note given with no overlapping. The 'i' vowel was chosen for the high tongue position and the benefits this has to singing. The external sound was created using a vocal synthesiser patch used in Vocal Tract Organ²² based on the software Pure Data.²³ This is used in the form of a button test (see Figure 1). Participants selected pitch by clicking on numbered buttons, each of which played the test sound at a different pitch, and their task was to select the button which played the note that they believe was closest to the target pitch. The subjects were allowed to repeat any stage of the button selection and compare it back to the target note until they were satisfied with their answer. However, they were not able to overlap the notes to make the comparison.

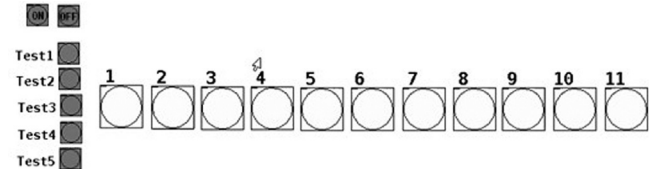


FIGURE 1. Button test screen shot from Pura Data²³ used for test 1.

The white buttons running horizontally numbered 1-11 are the variations of the target note set 10 cents apart ascending in pitch from left to right.

The sung notes are analysed using Praat¹⁹ for their f_o (fundamental frequency) which is condensed into an average for the duration of the note and then plotted into a chart.

Test two is designed to measure the effect of occlusion of the ear on tuning while singing and whether occlusion changes the perception of the pitch of an individual's voice to their overall tuning accuracy whilst singing a familiar song. The addition of occlusion adds the potential to attenuate own voice perception compared to air conduction which links back to the hypothesis: 'Is there a measurable psychoacoustic difference in how we perceive pitch via air conduction and bone and muscle conduction. This test is to test pitch matching in a 'real world' scenario where they will sing the familiar song 'Happy Birthday'. The findings will show whether there is a measurable difference in how pitch

is perceived via bone and muscle conduction and air conduction.

The participants were asked to sing the song ‘Happy Birthday’ three times starting on a given starting note appropriate for their voice fach or vocal range.

The first time, subjects sang ‘Happy Birthday’ normally with no occlusion, the second time they sang it with a pair of headphones in or on (for in-ear bud style headphones or over/ on ear headphones) to provide air-borne sound occlusion. The third time it is sung wearing headphones, connected to the device playing the instructional track, with white noise playing through the headphones to disrupt anything they might still be hearing via air conduction. The same pair of headphones was used for both versions of ‘Happy Birthday’. The volume of the white noise was set to be at a comfortable level on an individual basis agreed by the participant.

Each note from the song is analysed using Praat¹⁹ instruction ‘pitch read’ which generates a measurement of f_o for every 100th of a second for the duration of the note which gives a measure of the time point on the recording and the corresponding f_o .

As per test 1, an f_o average is taken for each sung note in cents deviation from the starting note. An overall average is calculated and plotted (see Figure 2). Each note of the 25 for the song ‘Happy Birthday’ is plotted on the X-axis of Figure 2 where the accuracy of the following notes is quantified in cents as the variation from the equal tempered tuned version of the target note. A variation of 100 cents would be a semitone, and an octave variation would be 1200 cents and a variation. of 1150 cents as the sung response to an octave jump, would indicate that the singer was 50 cents (or a quarter tone) flat in pitch.

Calculation of results

The difference in cents (‘The cent is defined as one-hundredth of an equal-tempered semitone, which is equivalent to one twelve-hundredth of an octave since there are 12 semitones to the octave’); is given in Equation 1¹² where F1 is the variable data point and F2 is the fixed-point average is:

$$C = 3986.3137 \times \log < ce : inf > 10 < /ce : inf > \frac{F1}{F2} \cdot 12 \quad (1)$$

The results from the vocal attempt button test are calculated to show the error¹⁹ from the target note in cents. An average of each f_o Hz is then taken for the duration of the note where a f_o readout is taken every 100th of a second; the average error for the note is then plotted on a chart (Figure 3).

The averages for all five notes in the voice and the button test of test one, and all 25 unique notes of ‘Happy Birthday’ for all three versions are plotted into charts to generate visual empirical data to demonstrate the comparative accuracy in test one and the overall melodic accuracy of test two.

For test 1, Equation 1 is used to calculate the distance in cents between the sung note and the target note and likewise with the button test.

The pitch drift is measured using the first sung note as the reference point rather than a target note This will show each singer’s ability to stay in tune relative to themselves. This is measured by interval accuracy, (as opposed to staying in tune with the exact Hz of the song ‘Happy Birthday’). Interval accuracy is quantified by calculating the frequency ratio between intervals in cents (e.g., ‘Ha-’ to ‘-ppy’ of Happy Birthday) using Equation 1 and repeated for each frequency ratio for all the notes in ‘Happy Birthday’. This enables inter-note musical intervals to be plotted as shown in Figure 3 when there is an octave leap, the expectation in to see an interval of 1200 cents or 12 semitones. This is done by subtracting the number of cents of the musical interval from the average of the sung data.

The results of the two tests are expressed in cents to demonstrate a uniform comparison of pitch accuracy across different voice types.

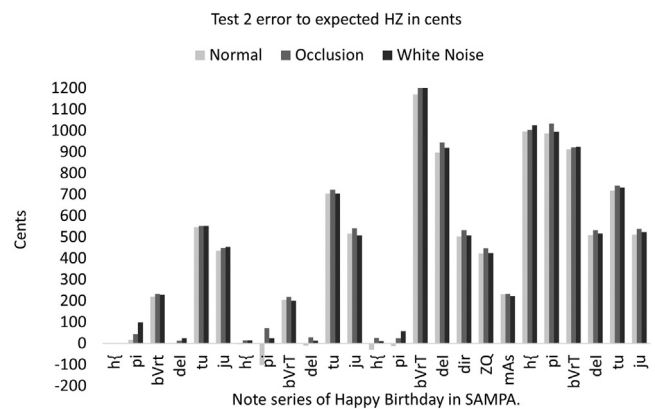


FIGURE 2. Plot of 3 versions of happy birthday showing the intervals in cents across the 25-note series shown in SAMPA phonetic alphabet (See Table 1).

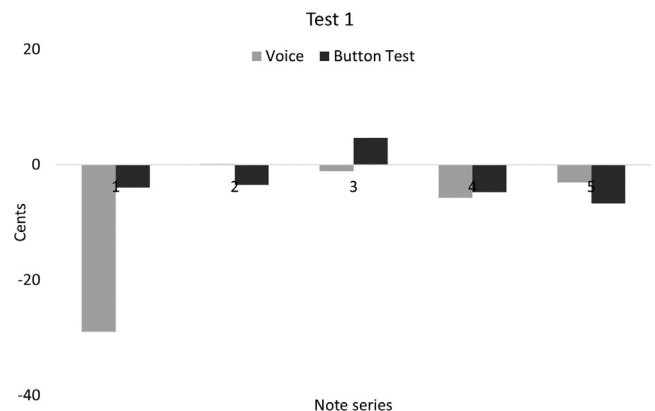


FIGURE 3. Plotted results of test 1 showing both vocal and button test attempts across the 5 notes of a single subject. This shows the error in cents from 0 on the X axis which is adjusted to represent each different pitch 1 -5. The adjustment is so that 0 on the X axis is representing 0 change in pitch for each note.

TABLE 1. The 24 Consonants and 20 Vowel Sounds in English With Their SAMPA Symbols (Sym.), an Example Word (Word) and Its SAMPA²⁴ Transcription (Trans.). The Voice, Place and Manner Descriptions are Listed for the Consonants.²⁵

English Consonants						English Vowels		
Sym.	Word	Trans.	Voice	Place	Manner	Sym.	Word	Trans.
p	Pun	/r@Up/	V-	Bilabial	Plosive	i	Neap	/nip/
b	Buoy	/bOI/	V+	Bilabial	Plosive	l	Jib	/dZlB/
t	Tide	/tald/	V-	Alveolar	Plosive	E	Red	/rEd/
d	Deck	/dEk/	V+	Alveolar	Plosive	{	Anchor	/ʔNk@/
k	Cabin	/k{bIn/	V-	Velar	Plosive	A	Hard	/hAd/
g	Galley	/g{li/	V+	Velar	Plosive	Q	Locker	/lQk@/
T	Thwart	/TwOt/	V-	Dental	Fricative	O	Port	/pOt/
D	Weather	/wED@/	V+	Dental	Fricative	U	Foot	/fUt/
f	Fog	/fQg/	V-	Labio-dental	Fricative	u	Food	/fud/
v	Vang	/v{N/	V+	Labio-dental	Fricative	V	Rudder	/rVd@/
s	Sea	/si/	V-	Alveolar	Fricative	3	Stern	/sT3n/
z	Zenith	/zEnIT/	V+	Alveolar	Fricative	@	Tiller	/tIl@/
S	Ship	/Sip/	V-	Palato-alveolar	Fricative	el	Weigh	/wel/
Z	Treasure	/trEZ@/	V+	Palato-alveolar	Fricative	al	Light	/alt/
h	Heeling	/hiIlN/	V-	Glottal	Fricative	OI	Oilskin	/OIIskIn/
tS	Chain	/tSelN/	V-	Palato-alveolar	Affricate	@U	Row	/r@U/
dZ	Jibe	/dZalB/	V+	Palato-alveolar	Affricate	aU	Bow	/baU/
m	Mast	/mAst/	V+	Bilabial	Nasal	l@	Pier	/pl@/
n	Main	/melN/	V+	Alveolar	Nasal	E@	Fare	/fE@/
N	Rigging	/rlgIN/	V+	Velar	Nasal	U@	Fuel	/fU@l/
w	Winch	/wIntS/	V+	Bilabial	Semi-vowel			
r	Rain	/rein/	V+	Alveolar	Semi-vowel			
l	Lee	/li/	V+	Alveolar	Semi-vowel			
j	Yacht	/jQt/	V+	Palatal	Semi-vowel			

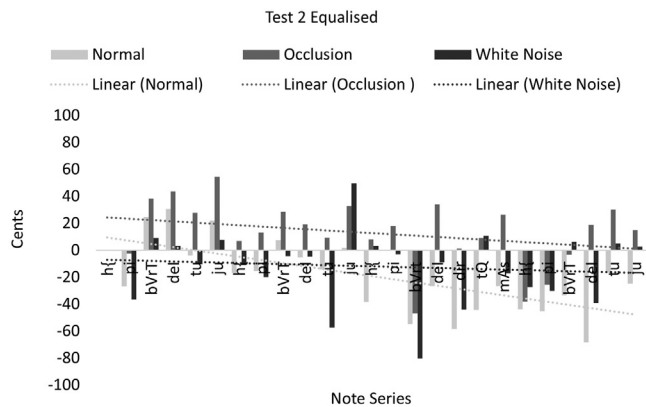


FIGURE 4. Results of chart 2 equalised to only show error from target pitch. The Linear is showing pitch drift.

RESULTS

The results of Table 2 show an overall average error from the target note in cents for the first test showing the vocal pitch recreation on the left and the button test pitch matching on the right. The averages show an overall average for all participants as well as a breakdown of the averages across the different gender groups and age groups which took part in the study.

TABLE 2. Results from Test 1.

Test 1 Results Voice	Test 1 Results Button Test
Overall average -8.23	Overall average -0.32
Male average -7.48	Male average 4.83
Female Average -8.99	Female average -4.19
21-30 Average -10.49	21-30 Average 2.29
31-40 Average -7.40	31-40 Average -0.84
41-50 Average 6.72	41-50 Average 12.24
51-60 Average -11.01	51-60 Average -4.17
61+ Average -12.68	61+ Average 1.76

TABLE 3.
Demographic Break Down of Participants Taken from a Study Group which Included 8 Men and 8 Women. The age Groups are Shown in Table 2.

Age	21-30	31-40	41-50	51-60	61+
	6.00	3.00	2.00	2.00	3.00
Gender					
F total	8.00				
M total	8.00				

The ratio of male to female participants was 50:50 which is shown in Table 3. Table 4 shows the average error in cents for the 3 different attempts of 'Happy Birthday' in test 2.

TABLE 4.
Results from Test 2.

Overall Average	Overall Average	Overall Average
23.62	39.46	10.94
Male average	Male average	Male average
21.00	35.40	12.21
Female average	Female average	Female average
26.23	42.78	9.66
Normal	Occlusion	White Noise
21-30 Average	21-30 Average	21-30 Average
17.34	26.65	19.17
31-40 Average	31-40 Average	31-40 Average
10.43	19.25	8.28
41-50 Average	41-50 Average	41-50 Average
15.41	55.31	2.36
51-60 Average	51-60 Average	51-60 Average
68.74	100.84	-2.08
61+ Average	61+ Average	61+ Average
24.75	33.79	11.53

The average error, or pitch drift (these are normally small deviations in pitch which cause the singer to drift off key and away from the target notes), is shown by +/- cents broken down into gender and age group as well as the overall trend for all participants.

The overall average error from the target pitch which demonstrates aural pitch matching using the button test (shown in Table 2) was -0.32 cents which as a total average suggests good pitch matching capabilities. An example of this can be seen in Figure 4. The range of the individuals average was from -18.84 cents (F4) to +27.61 cents (M1). The vocal pitch matching where the participant sang back the 5 pitches was on average -8.2 cents which is not as accurate as the aural pitch matching with matching the same 5 pitches with the button test which was -0.32 cents on average. The range for the individual average for this test was from -35.56 cents (M6) to +10.65 cents (M4) (see Table 5). Male subjects were, on average, -7.5 cents in the vocal pitch matching and -4.8 cents for the button test and female subjects were -8.9 cents on average when matching pitch with their voices and -4.2 of cents for the button test. The most accurate age group for vocal pitch matching were the 41's to 50's with +6.7 cents on average. The most accurate group for the button test was the 31-40's with -0.84 cents which is a smaller margin to the target pitch than the best vocal pitch matching. Given that it was found that trained singers couldn't discriminate better than 20 cents⁴ the overall trend of pitch drift would still be classed as in tune based of previous definitions. It should be noted that the pitch accuracy is based in equal temperament which is the same tuning system a piano would use. This has been used because often singers are singing in reference to an instrument to give them their pitch or are accompanied by an instrument(s) which is also in equal temperament.

TABLE 5.
Average Pitch Accuracy Across Both Tests for Each Individual Participants as Well as Detailing Their Age, Gender.

	Test 1 Results Voice	Test 1 Results Button Test	Test 2 Normal Average Pitch Drift	Test 2 Occlusion Average Pitch Drift	Test 2 White Noise Average Pitch Drift	Gender	Age
F1	-7.74	-2.84	-19.06	12.77	-11.83	F	31-40
F2	-7.74	-2.84	28.26	15.47	23.57	F	31-40
F3	-9.52	-7.31	1.89	31.10	23.71	F	21-30
F4	-18.51	-18.84	33.80	7.74	-2.13	F	61+
M1	-6.28	27.61	18.12	1.89	40.00	M	61+
M2	19.62	23.61	34.20	39.89	-5.38	M	41-50
M3	-23.09	3.61	-21.54	-3.25	32.70	M	21-30
F5	-6.19	0.86	-3.39	70.73	10.09	F	41-50
F6	-6.71	3.16	22.08	29.51	13.09	F	31-40
F7	-7.74	-2.84	65.04	75.28	24.62	F	21-30
M4	10.65	1.61	5.19	-59.51	10.07	M	21-30
M5	2.32	-0.39	35.50	60.02	34.65	M	21-30
M6	-35.56	-8.39	17.95	56.26	-10.73	M	21-30
M7	-14.28	-5.50	56.22	96.19	-0.34	M	51-60
F8	-7.74	-2.84	81.25	105.49	-3.81	F	51-60
M8	-13.24	-3.50	22.33	91.73	-3.28	M	61+

DISCUSSION

As this study was conducted remotely, using a range of microphones and devices for the recordings, a follow up study was conducted to investigate the effect the variable factors had on f_o . Three different recordings of the same source audio were taken simultaneously using a mobile phone, a PC using a Rode NT55 microphone, and an Ipad recorded over Zoom.¹⁸ This equipment test recorded a piano playing 5 notes as in test 1.

TABLE 6.
Results from the audio recording equipment test measured in Hz.

	Phone	P.C	Ipad
Pitch Number			
1	165.38	165.32	165.39
2	220.43	220.40	220.40
3	278.27	278.35	278.23
4	330.90	330.77	330.79
5	370.68	370.61	370.51

Each note in Table 6 was sampled over 10 cycles of the notes sound wave from peak to peak to get the same number of cycles for each result.

TABLE 7.
A Comparison of Table 6 Results in Cents.

Pitch Number	PC to Phone	Ipad to Phone	Ipad to PC
1	0.61	0.19	0.80
2	0.23	0.27	-0.04
3	-0.56	0.21	-0.77
4	0.66	0.56	0.09
5	0.34	0.81	-0.47

The results from Table 7 show that there is less than a cent difference between the three recording devices in this test. The same calculation of cents is used from the calculation of results section. This means that there is less than a $1/100$ of a semitone difference between the recording devices which is not a significant difference which would have impacted the results of this study.

It could be argued that these findings show that there is a marginal difference between vocal pitch matching and matching using aural perception in which aural skills are more accurate. This could be due to the physical ability to match pitch and vocal coordination rather than an inaccuracy to assess the pitch of one's own voice accurately. These findings support the findings of Loui et al.⁵ where perceptual abilities reflected by the pitch production. This also supports the results of Watts et al.¹³ where more than 40% of pitch

matching accuracy was due to pitch discrimination error. The purpose of test 1 was to ascertain if the participant could match pitch both aurally and vocally. The results show there is a small difference, but the aural perception and recreation of pitch is more accurate than that of the vocal pitch matching.

The difference in pitch accuracy between aural and vocal attempts could be down to vocal coordination rather than misjudgement of pitch as they hear it in the voice.

Test 2 is to measure the pitch drift in a melodic setting, comparing 3 different attempts of singing the familiar tune 'Happy Birthday', first time normally, second time with occlusion (by wearing headphones) and third time with white noise (by wearing headphones with white noise playing through them while singing). The results in Table 4 show that in all total averages for all 3 versions of 'Happy Birthday' drifted sharp. This could be a conscious drift in an attempt to stay in tune and not let the pitch drift flat as the pretext to the study was assessment of pitch perception. However, the version which showed the highest accuracy, and the least overall average drift was the white noise attempt.

The use of occlusion changing pitch drift is based on a studio method of masking the singer's ears in some form to stop pupils self-assessing their own singing to improve pitch accuracy and quality of sound. The outcome of it positively affecting pitch drift in a melodic setting could be used in a studio and teaching setting to help with this issue.

There is no obvious trend regarding age or gender in the results. A larger study would be required to assess if there is a trend based on demographic. Of all the participants who chose to answer the question rating the attempts of 'Happy Birthday' only 1 said they found the attempt with white noise to be their best and 3 said that occlusion was their best attempt whereas all other participants said the first attempt with no occlusion or white noise was the best, when in fact the data suggests that it was the second most accurate overall.

CONCLUSION

The results given in this paper show that on average within the sample group, there is a small difference in pitch matching abilities using one's voice and matching two external sounds, but it is noted that the latter of these was the more accurate. A larger study group would give more confidence in the results from this test. In addition, the differences in the results could be due to vocal coordination and the ability to achieve the desired pitch.

The second test shows that on average, pitch drift is reduced with the introduction of occlusion to the ear canal.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank all the subjects for taking part in these experiments especially given the restriction during the COVID-19 pandemic.

APPENDIX 1

Appendix 1. This shows the calibration of the piano notes given in the case of an Alto voice demonstrating the target pitch error from exact f_0 in cents.

APPENDIX 2

Appendix 2. This is the questionnaire that the participants completed for the study (the questions regarding the study were completed once they had completed the tests).

REFERENCES

1. Fant G. *Acoustic Theory of Speech Production*. Vol 2. Mouton; 1970.
2. Titze IR, Martin DW. Principles of voice production. *J Acoust Soc Am*. 1998;104:1148. <https://doi.org/10.1121/1.424266>.
3. van der Woerd B. *Voice Analysis with Iphones: A Low Cost Experimental Solution*. ProQuest Dissertations Publishing; 2019. <https://search.proquest.com/docview/2714870337>.
4. Howard DM, Angus JAS. A comparison between singing pitching strategies of 8 to 11 year olds and trained adult singers. *Logopedics Phoniatrics Vocol*. 1997;22:169–176. <https://doi.org/10.3109/14015439709075331>.
5. Loui P, Demorest SM, Pfordresher PQ, et al. Neurological and developmental approaches to poor pitch perception and production. *Ann N Y Acad Sci*. 2015;1337:263–271. <https://doi.org/10.1111/nyas.12623>.
6. v. Békésy G. The structure of the middle ear and the hearing of one's own voice by bone conduction. *J Acoust Soc Am*. 1949;21:217–232. <https://doi.org/10.1121/1.1906501>.
7. Békésy G. *Experiments in Hearing*. Acoustical Soc. of America; 1960.
8. von Békésy G, Peake WT. Experiments in hearing. *J Acoust Soc Am*. 1990;88:2905. <https://doi.org/10.1121/1.399656>.
9. Fourcin A. Hearing and Singing. Chapman J, Howard DM, eds. Fourth edition. ed. San Diego, CA: Plural Publishing Inc; 2021.
10. Sundberg J. *The Science of the Singing Voice*. Northern Illinois Univ. Pr; 1987.
11. Cobes CJ. The conditioning of a pitch response using uncertain singers. *Bull Counc Res Music Educ*. 1972;28–29. <https://www.jstor.org/stable/40317272>.
12. Howard DM, Angus JAS. *Acoustics and Psychoacoustics*. 5th ed. Routledge; 2017.
13. Watts C, Murphy J, Barnes-Burroughs K. Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with nontalented singing voices in conditions of varying feedback. *J Voice*. 2003;17:185. [https://doi.org/10.1016/S0892-1997\(03\)00023-7](https://doi.org/10.1016/S0892-1997(03)00023-7).
14. v. Békésy G. Vibration of the head in a sound field and its role in hearing by bone conduction. *J Acoust Soc Am*. 1948;20:749–760. <https://doi.org/10.1121/1.1906433>.
15. Welch GF. *Improvability of poor pitch singing experiments in feedback [dissertation]*. Institute of Education, University of London; 1983.
16. Haynes B, Cooke P. Pitch. *www.OxfordMusicOnline.com* Web site. <https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000040883?rsk=JkVeZ>. Updated 2001. Accessed 27/01, 2022.
17. Howard DM, Hunter EJ. *Perceptual Features in Singing*. *Perceptual Features in Singing*. Oxford University Press; 2019.
18. Eric Yuan. Zoom. 2011.
19. Boersma P, Praat. 1991;6.1.49.
20. Uloza V, Padervinskis E, Vegiene A, et al. Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. *Eur Arch Otorhinolaryngol*. 2015;272:3391–3399. <https://doi.org/10.1007/s00405-015-3708-4>.
21. Titze IR, Winholtz WS. Effect of microphone type and placement on voice perturbation measurements. *J Speech Hear Res*. 1993;36:1177–1190. <https://doi.org/10.1044/jshr.3606.1177>.
22. Howard DM. The vocal tract organ: a new musical instrument using 3-D printed vocal tracts. *J Voice*. 2018;32:660–667. <https://doi.org/10.1016/j.jvoice.2017.09.014>.
23. Puckette M. Pure Data. 1996;0.52-2.
24. Wells JC. Computer-coded Phonemic notation of individual languages of the European community. *J Int Phonetic Assoc*. 1989;19:31–54. <https://doi.org/10.1017/S0025100300005892>.
25. Plural Publishing I, Murphy D. *Voice Science, Acoustics, and Recording*. Plural Publishing, Incorporated; 2007.