

COVER PAGE

Document downloaded by @DAEL

Wed Nov 25 08:02:48 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Determining the Just Noticeable Difference in Timbre Through Spectral Morphing: A Trombone Example, Sandra Carral, Acta Acustica **vol. 97** (Number 3), 2011, pp. 466-476

DOI
https://doi.org/10.3813/AAA.918427

Acta Acustica united with Acustica
Vol. 97 (2011) 466 – 476

DOI 10.3813/AAA.918427

Determining the Just Noticeable Difference in Timbre Through Spectral Morphing: A Trombone Example

Sandra Carral*
School of Physics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom.
carral@mdw.ac.at

Summary
The produced sound and corresponding spectrum of musical instruments are usually affected to some degree when small variations in the parameters of the instrument are made. However, the question remains, of how much a certain parameter has to vary in order for the instrument to produce a difference in sound that can be perceived. The subject of this paper is to determine the Just Noticeable Difference between pairs of sounds corresponding to morphs (interpolations) between sounds of a trombone played with two different mouthpieces. This was done with the aid of a 2-alternative forced choice psychoacoustical test, where pairs of synthesised signals with varying degrees of spectral difference (generated through spectral morphing) were presented to subjects in randomised order.

PACS no. 43.66.Jh, 43.75.Fg

1. Introduction

Some of the research in timbre of musical sounds has mainly focused on the difference in timbre between several musical instruments. Many authors have done psychoacoustical tests and through multidimensional scaling (MDS) techniques have generated spaces with several dimensions where different instruments are placed. The placement is done so that the closer they are to one another in the space, the more similar their timbres are perceived in psychoacoustical tests. With this information, attempts are made to correlate the dimensions obtained through MDS to acoustical attributes of the sound, which can be measured (see for instance [1], [2], [3]).

Some authors have studied timbre recognition by recording instrument sounds, and removing some of their properties, for instance, attack or vibrato (see for instance [4], [5], [6]). A more sophisticated method has been of analysis-synthesis of musical sounds, the resynthesis being made with various degrees of simplifications, in order to study the perceptual effects of the removed components (see for instance [7], [8], [9], [10], [11]).

However, little has been studied on the differences within a particular instrument. For example, a woodwind instrument can be played with reeds of different materials and properties; or a flute can be played with different head joints, etc. It is expected that the spectrum of the radiated sound of the instrument will indeed be affected by these changes. As a matter of fact, most musicians would agree that these changes affect the sound of the instrument, all other things kept equal. Finding out just how different two sounds have to be in order for people to be able to hear a difference could help in determining whether small changes in the parameters of a particular instrument result in a sound that is perceivably different from the instrument in its original state, without having to resort to a time consuming psychoacoustical test.

In this paper, the question of how much difference in the spectral content of two sounds can be perceived is addressed. This was done through spectral morphing of two sounds recorded from a trombone played with two different mouthpieces. Pairs of sounds with varying degrees of spectral difference (intermediate steps in the morphing) were presented in randomised order in a two alternative forced choice (2AFC) psychoacoustical test to determine the Just Noticeable Difference (JND). Since timbre is a multidimensional quantity, it is at first glance not straight forward to state what the difference is between two distinguishable spectra. The parameters that were measured in this paper were: 1. the difference in harmonic amplitudes of the spectra in dB and its mean absolute value, and 2. the difference in the normalised spectral centroid vs time and its mean absolute value.

This paper is organised as follows: The rest of this section presents work that has been done on JND of harmonic

Received 22 April 2010,
accepted 28 February 2011.

* Present affiliation: Institute of Musical Acoustics, University of Music and performing Arts, Vienna, Anton von Webern Platz 1, 1030 Vienna, Austria

© S. Hirzel Verlag · EAA

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

signals, and perceptual timbre differences within the same instrument. Section 2 describes how the signals from the trombone sounds were produced and recorded. The differences in the parameters of, in this case, the mouthpiece, will in turn produce different spectra. The differences in harmonic content between these signals are studied.

Section 3 describes the differences found in the spectra of the recorded sounds. The next step was to synthesise the recorded sounds, with the purpose of being able to modify the amplitudes of each harmonic independently from one another, and yet when not modified, be indistinguishable from the originals. Such method is described in section 4, highlighting which components of the original signals are considered to be crucial to making the synthesised signals indistinguishable from the originals.

Section 5 presents a psychoacoustical test, where pairs of sounds with controlled differences in harmonic content were presented to subjects in randomised order, so as to be able to determine the JND.

1.1. Related work on timbre

To the author's knowledge, there is only one study where the JND of timbre of sounds from the same instrument was investigated: Coltman [12] resynthesised sounds from different notes of a flute, changed the amplitude of one harmonic at a time, and with these signals he made psychoacoustical tests to find the JND. He found that the changes had to be between 0.7 and 3 dB depending on which harmonic was changed, its amplitude relative to neighbouring harmonics, and whether the amplitude was increased or decreased. Finally, he analysed the differences in harmonic content of a flute played with two different head joints, and with the results he had obtained concluded that the differences should be perceptible.

Another study involving the determination of the JND in timbre, specifically in spectral centroid, was done by Kendall and Carterette [13]: They studied the psychoacoustical effect of changing the spectral centroid using artificial signals. They found that the just noticeable difference in normalised (dimensionless) spectral centroid was between 0.11 (for musicians) and 0.15 (for non musicians).

Nishimura *et al.* investigated the perception of timbre of flute sounds [11]: They recorded flute sounds, and by analysis-synthesis they altered the amplitude of the random amplitude and frequency fluctuations above the vibrato frequency (from 13 Hz). They performed a psychoacoustical test where people were presented pairs of sounds with different degrees of frequency and amplitude fluctuations, and the subjects were asked to state which of the two sounds they preferred. They concluded that listener's preference was affected by the degree of random fluctuations that occur above 13 Hz.

Wright and Campbell [14] studied the differences in the sound of trumpets and cornettos when played with different mouthpieces, and did psychoacoustical tests to confirm that these were perceptible. They found that even when the differences in the spectra of the sounds were very small, the listeners could still distinguish between different mouthpieces. Then with the aid of synthesised sounds, they tried to find which were the features of the sound to which the listeners were sensitive. They concluded that the amplitudes of partials relative to the noise floor are more important than the absolute amplitudes. Similarly, the high frequency components are used as a "reference level" from which changes at lower frequencies are calibrated.

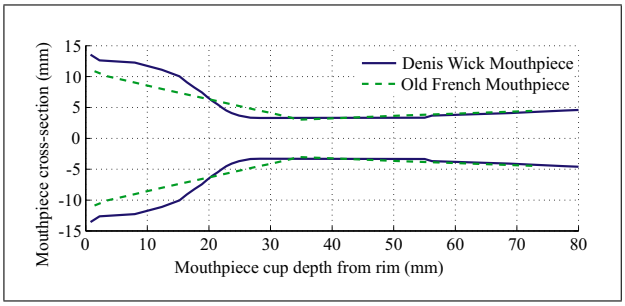


Figure 1. Bore profile of the mouthpieces used for the experiments.

2. Recordings of trombone sounds using two different mouthpieces

A semi-professional trombone player was asked to play a tenor trombone model King 2B with two different mouthpieces: a standard Denis Wick model 6BS and a nineteenth century French mouthpiece from the Edinburgh University Collection of Historic Musical Instruments, catalogue number 3726, subsequently referred to as the "Old French mouthpiece".

The bore profile of both mouthpieces is shown in Figure 1. For simplicity purposes, in the following sections, references to "Denis Wick Mouthpiece" or "Old French Mouthpiece" will refer to the trombone played with one or the other mouthpiece. A preliminary psychoacoustical test showed that the sounds produced by this trombone using the two mouthpieces were easily distinguishable.

To minimise reflections that could change the shape of the spectrum, the recordings were made in an anechoic chamber, using an Audio-Technica omnidirectional condenser microphone model ATM10a fixed on the trombone bell with a harness at a distance of 1 bell diameter, as this would produce a spectrum like that of having averaged several spectra taken at different places in a reverberant room [15] p. 347, [16] p. 421.

The microphone was connected to a TASCAM mixer model M–08 and the level of the signals was adjusted to be as high as possible without saturating. This adjustment was done once at the beginning of the measurements. The output of the mixer was then connected to a TASCAM DAT recorder model DA-20mkII, which recorded at a sampling frequency of $f_s = 44.1$ kHz.

A DAWE sound level meter model D-1422 was placed about one meter away from the bell of the instrument, where the player could clearly see the value measured. It measured approximately 80 dB SPL when the player

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

was playing at what he considered a piano dynamic level. Hence, he was asked to play in such a way that the sound level meter would measure 80 dB SPL, and to play the notes steadily and without vibrato.

The player played first a few arpeggios to warm up, and then some steady notes. From these recordings, a $B_3^\flat$ (concert pitch) note from each of the two mouthpieces was selected. The original selected notes were played steadily for about 4 seconds. A signal of 0.5 s duration (22050 samples) was taken from each of the two notes.

3. Analysis of recorded signals, and comparison of spectra of signals from the two mouthpieces

Since the harmonic content of the signal from the trombone signal played with one mouthpiece was to be compared with that played with the other mouthpiece, it was important to make sure that the amplitudes of these two signals were the same. Otherwise, the differences in spectral content would be affected by both the difference in amplitude between the signals, and by the effect of the mouthpiece. Therefore, the RMS value of these signals was equalised.

The loudness level of the equalised signals was calculated using a MATLAB program written by Marek Dziubinski from the Technical University of Gdansk, Poland, based on a method developed by Glasberg and Moore [17]. The difference in loudness level between the equalised signals was less than one phon. The JND in intensity level (IL) is between 0.5 (for loud sounds) and 1.5 (for soft sounds) dB [18]. Although phons are not dB's, the difference in loudness of less than one phon found between these two sound samples was considered to be imperceptible to the human ear.

The 0.5 second signals were windowed (Hanning) and zero-padded so as to have 44100 samples. A Fast Fourier Transform (FFT) was taken for each signal. These spectra are shown in Figures 2 and 3.

Figure 4 shows the difference in amplitude (dB SPL) of each harmonic of these two spectra. In Figure 4, the height of the bars represents the extent to which the Denis Wick mouthpiece harmonic amplitude in dB exceeded that of the Old French mouthpiece. Negative values imply that the Old French mouthpiece harmonic was higher than that of the Denis Wick mouthpiece. It is worth noting that the FFT's that resulted in the spectra shown in Figures 2 and 3 were taken across the whole 0.5 duration of the windowed signals, hence the differences shown in Figure 4 are average differences, which means that time dependent variations in the spectral amplitudes are not taken into account.

It has been reported in the literature (see for instance [19], [20] and [10]) that time fluctuations of the spectrum are important for the perception of timbre. In the light of this, it was decided to do a time dependent frequency analysis on the unwindowed 0.5 second signals. This was done with the aid of the pitch synchronous phase-vocoder

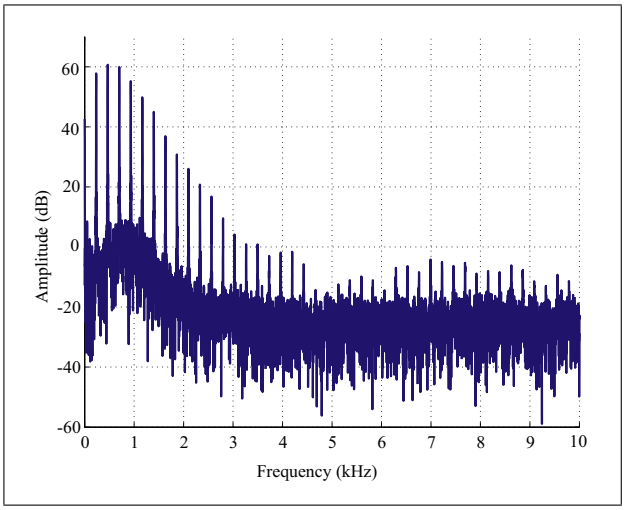


Figure 2. Spectrum of signal obtained from the trombone played with the Denis Wick mouthpiece.

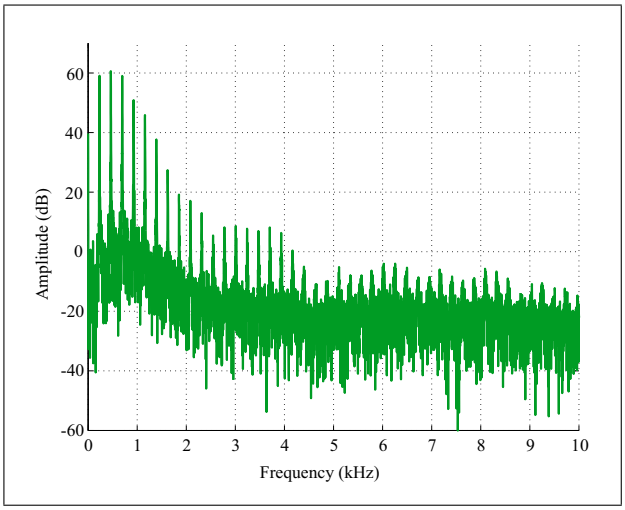


Figure 3. Spectrum of signal obtained from the trombone played with the Old French mouthpiece.

analyser that comes with the SNDAN package written by James W. Beauchamp. For details about how this program performs the analysis, the reader is referred to [21] and [22]. The program expects as input the fundamental frequency of the signal to be analysed, and is referred to as "analysis frequency" f_a. Since the played note was $B_3^\flat$, f_a was chosen to be 233.1 Hz, which is the nominal playing frequency of this pitch in an equally tempered scale with $A_4 = 440$ Hz. Since each analysis window is 2 periods long, and is advanced by half a period $\frac{1}{2f_a}$, the analysis frame duration is approximately 2.15 ms. The program delivers a file with information about the harmonic amplitudes A_k and frequency deviations Δf_k from the kth harmonic of the analysis frequency f_a at every time frame.

With this information, the composite weighted-average (CWA) frequency deviation in cents relative to the analysis frequency f_a at each analysis frame can be calculated [22]:

$$\Delta f_{cwa}(t) = 1200 \log_2 \left(\frac{f_a + \Delta f_c(t)}{f_a} \right), \qquad (1)$$

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

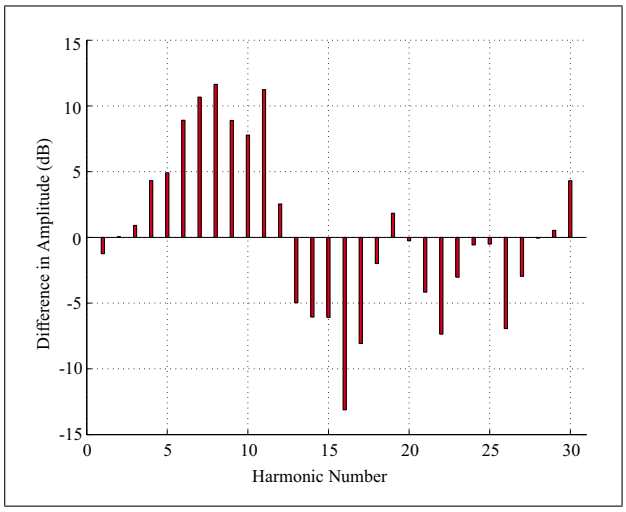


Figure 4. Spectral differences in harmonic content between Denis Wick and Old French mouthpieces. Positive dB values represent the excess of the Denis Wick mouthpiece over the Old French mouthpiece. All values are averages over the whole 0.5 second duration of the windowed signal.

where

$$\frac{\Delta f_c(t)}{f_a} = \frac{\sum_{k=1}^{5} \frac{A_k(t)\Delta f_k(t)}{k \cdot f_a}}{\sum_{k=1}^{5} A_k(t)} \tag{2}$$

is the normalised composite weighted-average frequency deviation, k is the harmonic number, $A_k(t)$ is the amplitude variation over time of kth harmonic and $\Delta f_k(t)$ is the frequency deviation over time of kth harmonic.

Likewise, the normalised spectral centroid (dimensionless) is calculated [21], [22],

$$NSC = \frac{\sum_{k=1}^{n} k\, A_k}{\sum_{k=1}^{n} A_k}, \tag{3}$$

where $n = f_s/(2f_a)$ is the maximum number of harmonics that a signal of fundamental frequency f_a sampled with a sampling frequency f_s can have.

Figures 5 and 6 show the CWA frequency deviation and normailised spectral centroid variations over time respectively. These curves are strictly speaking not continuous, rather each point corresponds to the calculation made within one analysis frame, but since the frame duration was about 2 ms it was chosen for clarity to connect the points.

The main differences found in the signals of these two mouthpieces are: 1. Harmonic amplitudes differ, as shown in Figure 4: these differences vary from 0 to 13 dB; and 2. the signals are not stationary, that is, they exhibit some slow modulation in both frequency (see Figure 5) and harmonic content (see Figure 6). This is not surprising, since even when a musically trained subject is asked to play steadily and without vibrato, this is not an easy task, mainly because musicians have been trained to produce beautiful sounds that have at least certain degree of vibrato.

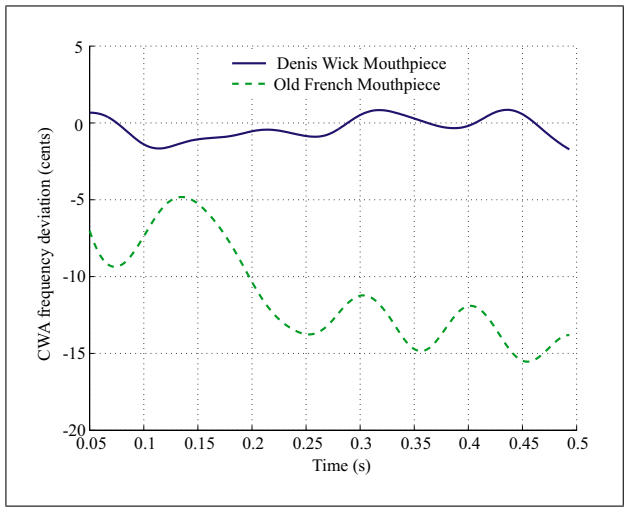


Figure 5. CWA frequency variations over time for each mouthpiece (0 cents corresponds to a frequency of 233.1 Hz).

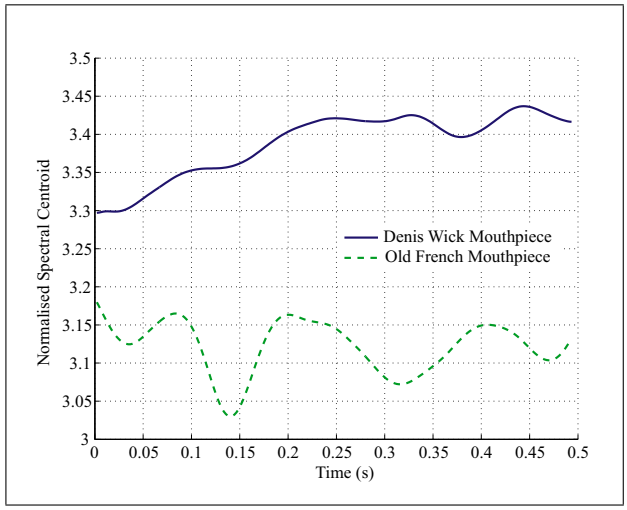


Figure 6. Normalised spectral centroid (dimensionless) variations over time for each mouthpiece.

Figure 5 shows that the note of the Denis Wick Mouthpiece was well in tune (within ± 10 cents relative to 233.1 Hz), whereas that of the Old French Mouthpiece was up to 15 cents flat. The average of the absolute difference between the normalised spectral centroid of the two curves shown in Figure 6 was calculated and found to be 0.26, the maximum difference being 0.34, and the minimum 0.09: Comparing these values with the difference threshold found by Kendall and Carterett e [13] (which were 0.11 for musicians and 0.15 for non musicians), this difference should be enough to be perceived.

4. Spectral Morphing

4.1. Overview

A very thorough timbre morphing algorithm was developed by the CERL sound group [23], [24]. Tellman *et al.* [23] define timbre morphing as

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

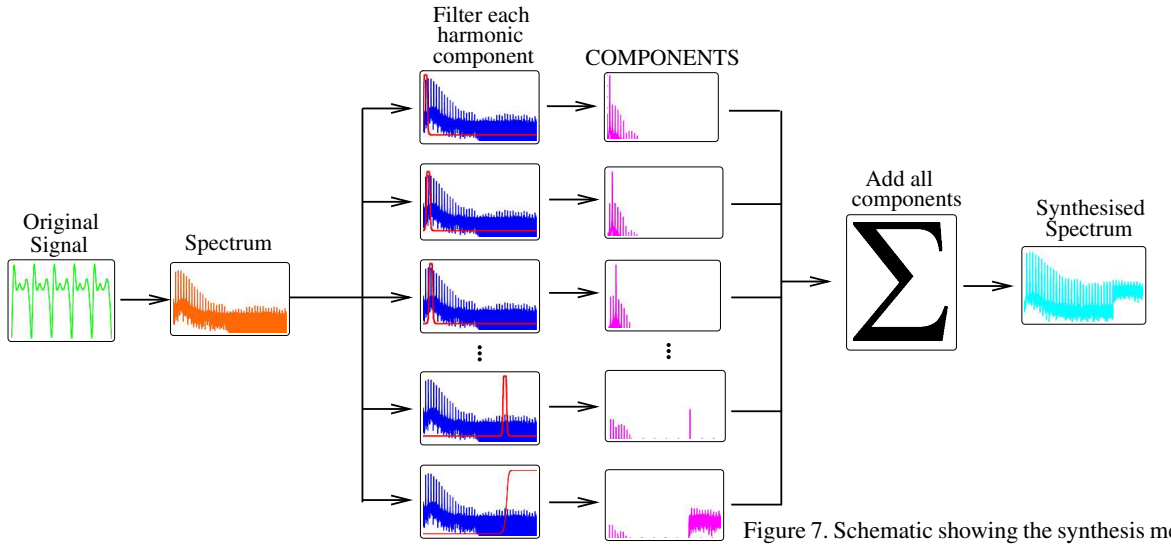


Figure 7. Schematic showing the synthesis method.

> "the process of combining two or more sounds to create a new sound with intermediate timbre and duration."

The algorithm they developed morphs attributes of the sound such as total duration, attack and decay time, vibrato as well as the relative amplitude of spectral components.

For the purposes of this paper the objective of using spectral morphing is to provide a smooth spectral transition between the sounds from the two trombone mouthpieces. A similar attempt, where only the spectra and the temporal variations of the sounds are morphed (the duration was equalised), was recently made by Horner *et al.* [25]. They studied the listeners' ability to discriminate between the original sound of a particular instrument and the sound morphed with another instrument at a morph extent. Their objective was to find to what extent a certain instrument has to be morphed with another instrument in order for people to tell that the two sounds are different. They found that the morphing discrimination depended on which two instruments were being morphed.

The difference between their study and the one presented here, is that rather than studying the discrimination of morphed sounds for several different instruments, this study is aimed at finding a spectral difference JND through spectral morphing between sounds played by a single instrument, albeit with two different mouthpieces.

The spectral morphing presented here is done by gradually reducing the differences in spectral content shown in Figure 4. The search for the spectral content JND is only possible if during the morphing process all other sound attributes (such as duration, attack and decay time, and the amplitude and frequency variations shown in Figures 5 and 6) are kept unchanged.

To do this, it is necessary to generate a synthesised version of one of the original signals, in order to be able to independently modify the spectral component amplitudes.

4.2. Synthesis method

This section explains how the trombone signal using the Denis Wick mouthpiece was resynthesised. The synthesised signal should have all the attributes that allow people to discriminate between two sounds. These attributes are: 1. harmonic amplitudes, 2. slow frequency and amplitude modulations and 3. high frequency aperiodic components.

The following procedure allowed the synthesised signal to include high frequency components such as noise, and retain the original slow frequency and normalised spectral centroid modulations shown in Figures 5 and 6.

The procedure to do this was:

1. the harmonics of the original signals were located, up to 7 kHz (30 harmonics);
2. each of these harmonics was filtered using a MATLAB band-pass FIR 4000th-order filter (window method), with a bandwidth of 2 Hz. This bandwidth would allow a frequency variation of ±7 cents on the fundamental;
3. all the information above the highest harmonic was filtered using a MATLAB high-pass FIR 4000th-order filter (window method);
4. the synthesised signal was generated by adding the output of filters (2) and (3).

A schematic of this synthesis method is shown in Figure 7. The filters used had a transient of around 90 ms, corresponding to 4000 samples. The selected portion taken from the original signal was extended to 26050 samples (by taking 4000 samples immediately previous to the original 22050 samples), and after filtering, the first 4000 samples of the filter transient were discarded. The samples that were kept (4001 to 26050) corresponded to the steady state of the filter. The 0.5 s duration of the sound was maintained.

By using this synthesis method, it was ensured that the following components of the sound were included in the synthesised signal: the high frequency aperiodic components were included as the output signal of the high-pass

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

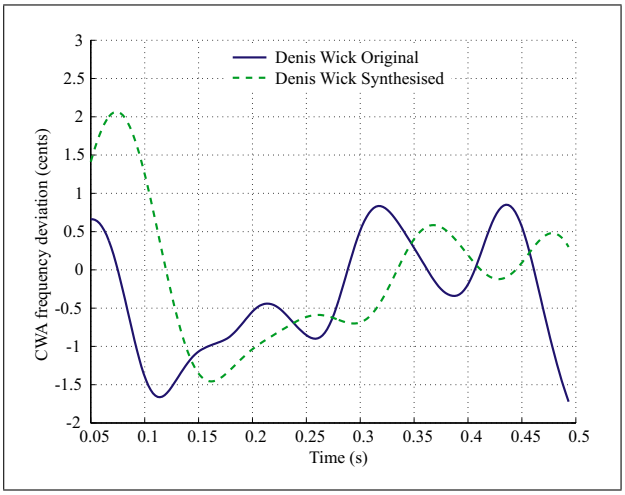


Figure 8. CWA frequency variation over time of original and synthesised versions from the Denis Wick mouthpiece (0 cents corresponds to a frequency of 233.1 Hz).

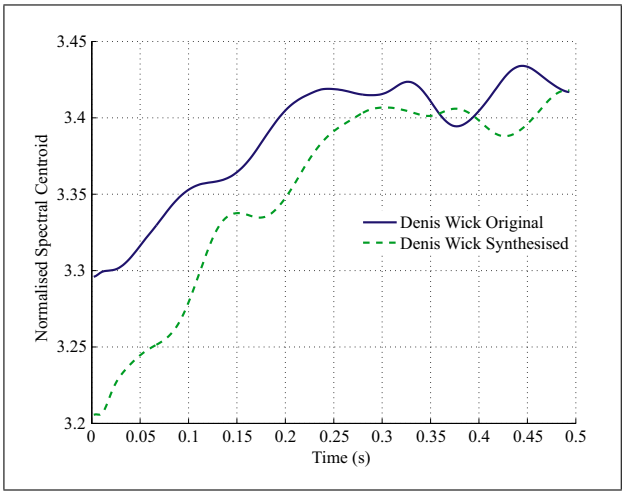


Figure 9. Normalised spectral centroid (dimensionless) variation over time of original and synthesised versions from the Denis Wick mouthpiece.

filter, and the slow frequency and amplitude modulations of the original signal were copied to the synthesised signals, because the harmonics of the original signal were filtered and then added unaltered to the synthesised signal.

A comparison between the spectra of the original vs the synthesised signal (not shown here) shows that the main difference is that the synthesised signal had a lower noise level in between harmonics. The band-pass filters used to filter out the harmonics were such that the noise level in between harmonics was reduced by around 40 dB SPL.

Figure 8 shows the CWA frequency variations over time of the original and synthesised signal. Likewise, Figure 9 shows the normalised spectral centroid variations over time. These two figures show that at least some of the slow frequency and amplitude modulations were kept in the synthesised version of the signal. From Figure 8 the mean of the absolute difference in CWA frequency between original and synthesised signals was 0.77 cents: The difference in CWA frequency between original and synthesised signals was well below the JND, which for complex tones is around 9 cents [15]. Likewise, from Figure 9, the mean of the absolute difference in spectral centroid was 0.04: The difference in spectral centroid was also below the JND reported in [13] (0.11 for musicians and 0.15 for non musicians).

The loudness level of the synthesised signal was calculated. The difference between the loudness level of the original signal and that of the synthesised signal was 0.3 phons.

In order to ensure that the synthesised signal was indistinguishable from the original, a 2AFC psychoacoustical test was made with 5 subjects, where pairs of signals (original-original, original-synthesised, synthesised-original and synthesised-synthesised) were presented in randomised order. After each pair was presented, subjects were asked to respond whether the signals in that pair were "THE SAME" or "DIFFERENT". There were 20 pairs in total, 10 same and 10 different. The results of this test

Table I. Contingency table showing the results of the psychoacoustical test to determine whether people could distinguish between original and synthesised signals from the Denis Wick mouthpiece.

		ANSWER	
		Same	Different
PLAYED	Same	42	8
	Different	37	13

are shown in Table I. A χ^2 test was performed, and since $p > 0.3$ it was concluded that the synthesised signal was indistinguishable from the original.

It is worth noting at this point that even though the slow temporal modulations of the synthesised signals were not identical to those of the original signals, the information that was kept in the synthesised signals was enough for them to be perceptually indistinguishable from the originals.

4.3. Spectral morphing between the signals from the two mouthpieces

The idea of spectral morphing is to gradually modify the spectral content of the Denis Wick mouthpiece tone until it is identical to that of the old French mouthpiece tone, but keeping the temporal variations of the Denis Wick mouthpiece tone. In this way, only the spectral content difference between the two tones is being evaluated, that come from having played with a different mouthpiece, and not the temporal variations that come from how the player played that particular note.

It was chosen to take 30 intermediate steps or morphing levels in the morphing, so that the sounds that resulted between any two contiguous levels were indistinguishable from one another.

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

The harmonic differences between the two mouthpieces that are shown in Figure 4 were divided into 30 equal steps. Since it was chosen to synthesise the signal from the Denis Wick mouthpiece, the morphing was done from the Denis Wick mouthpiece signal (reference signal) into the Old French mouthpiece signal (target signal). Each step in the morphing resulted in a new signal (morphed signal at a particular morphing level MS_l) where the harmonic amplitude differences between the reference signal and the new intermediate signal were increased by one step of size $\frac{1}{30}$ of what is shown in Figure 4 (likewise, the difference between the target signal and the new intermediate signal were reduced by the same amount). For instance, the first intermediate signal MS_1 resulted from modifying the harmonic amplitudes of the reference signal by $\frac{1}{30}$ of what is shown in Figure 4, the second one MS_2 by $\frac{2}{30}$, the third one MS_3 by $\frac{3}{30}$, until the last signal MS_{30} by $\frac{30}{30}$, at which point the spectral differences between the latter signal and the target signal were zero in all harmonics. This procedure can be summarised as

$$MS_l(t) = \left(\sum_{k=1}^{N} d_k^{l/s} \cdot H_k(t) \right) + HFN(t), \qquad (4)$$

where $MS_l(t)$ is the resulting morphed signal at a particular level l (in this case l is an integer from 1 to 30), $H_k(t)$ is the kth harmonic component of the reference signal (output of the kth bandpass filter from Figure 7), d_k is the difference factor that corresponds to the kth harmonic, s is the total number of morphing steps (in this case 30) and $HFN(t)$ is the high frequency noise components above 7 kHz (from the reference signal). The terms $MS_l(t)$, $H_k(t)$ and $HFN(t)$ are each a vector of samples in the time domain with $f_s = 44.1$ kHz and duration of 0.5 seconds.

The difference factor d_k for each harmonic component is obtained as follows: From Figure 4 the difference in dB $d_{k[dB]}$ for each harmonic component is taken, and converted from dB to pressure level as

$$d_k = 10^{d_{k[dB]}/20}. \qquad (5)$$

Neither the frequency nor the small amplitude variations of the harmonics were altered by the morphing, rather only each harmonic component from the reference signal was multiplied by a factor so as to alter its amplitude throughout the duration of the signal. Therefore, the last signal, whose harmonic amplitudes corresponded to those of the target signal, had nevertheless both the same fundamental frequency, and the frequency and amplitude fluctuations as the reference signal.

5. Just noticeable difference in timbre of trombone sounds

The just noticeable difference between two sounds is the smallest change in a stimulus that is perceivable by a human being [26]. It is a statistical quantity, therefore it is necessary to conduct many trials in order to determine the threshold. It usually indicates the difference that a listener detects on 50% of the trials.

The method used to find the JND in this paper has been referred to in the literature as the method of constant stimuli (see for instance [27], Appendix D) or fixed point (see for instance [28]). The advantage of this method is that a psychometric function can be obtained directly from the listeners' responses, which is a plot of the listeners' performance vs a value of the stimulus. The estimate of the JND can be obtained from a point on the psychometric function where the listeners' performance is e.g. 50%. This section describes such test, aiming to find a JND in spectral difference between the sound of the trombone played with the two mouthpieces described previously.

Another possible method that can be used to obtain the JND is referred to as "up-down" [29], "threshold-seeking" [28] or "adaptive" [30] method, where the following stimuli depends on the previous response: The test starts with a certain initial value of the stimulus variable. When the response is positive, the value of the stimulus variable is lowered. When the response is negative, the stimulus variable is increased. The threshold is then obtained by analysing the sequences of positive and negative responses. The main difference between the method of constant stimuli and the adaptive method, is that in the latter only the threshold point is obtained (as opposed to the whole psychometric function), since all the presented stimuli fluctuate around the threshold [28].

5.1. Description of psychoacoustical test to determine the JND

The 26 subjects used for this experiment were Physics students in the University of Edinburgh, who were asked to volunteer in a psychoacoustical test "that has to do with how humans perceive sound". It was determined by a short listening test that the reference signal and the 15th intermediate signal from section 4.3 MS_{15} were easily distinguishable. The signals chosen for testing were the first 15 intermediate steps out of 30 (signals MS_1 to MS_{15}). These will be henceforth referred to as test signals.

In this test, the reference signal was paired with the 15 test signals and presented to the subjects. Each pair consisted of the reference signal, and either the reference signal (two equal signals) or a test signal chosen randomly. The order of presentation of the two signals (that is, whether the first signals was reference or test) was also random. Every signal in all pairs had a duration of 0.5 s. A linear attack and decay, each of 50 ms duration was introduced to all signals. The test was made using a computer program, that presented the pairs in randomised order to the subjects.

Each pair was presented 20 times, giving a total of 320 presentations. Out of these 320 pairs, only 20 of them consisted of the reference signal played twice (control pair). The subjects were unaware of this. The sounds were presented through a pair of Sennheiser headphones model HD433. The volume was preset at a comfortable level at

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

the start of each test. The test was performed in a sound proof laboratory. Every pair presented played one sound, followed by a 0.5 pause, followed by the second sound. The subjects were allowed to listen to every pair as many times as desired before answering. There was a training session prior to the test, in which all the 16 pairs were presented once in randomised order, to allow the subject to familiarise him/herself with the procedure. The instructions to the subjects were:

> In this test you'll hear pairs of sounds, and you'll be asked to state whether they're
>
> THE SAME
>
> or
>
> DIFFERENT.
>
> You'll be able to hear each pair as many times as you like before answering. You'll state your answers by pressing the buttons labelled THE SAME or DIFFERENT, and you'll be able to hear the pair again by pressing the button labelled PLAY AGAIN.
>
> It will start with a short training session, so that you familiarise yourself with the procedure. During the training session your answers will not be recorded. To start the training session press START SESSION

After the training session the instructions presented on the screen were:

> TRAINING SESSION FINISHED
>
> Now you can start the Main Test by pressing the button labelled START SESSION again.

Once the main session started after pressing the START SESSION button, the screen included a progress bar, which helped the subject to know at which stage of the test he/she was at any particular time. Subjects were allowed to take a short pause in the middle of the test, if so they wished.

The program saved in a file the number of times that the subject answered DIFFERENT for each pair, resulting in a one row vector with 16 columns, each column corresponding to the degree of difference between the reference and test signals.

5.2. Results of the psychoacoustical test

The vectors for all subjects were added, resulting in the number of times that all subjects answered "DIFFERENT" for all 16 degrees of difference. This vector was processed to give the proportion of DIFFERENT responses q_i for each degree of difference i, which varies from $i = 0$ (no difference) to $i = 15$ (maximum difference). A plot of q_i discrimination vs i would show on the y axis the proportion of DIFFERENT responses q_i for the degree of difference i, and it is expected that q_i will increase as i increases. Such a plot is shown in Figure 10, and as expected, q_i increases as i increases. Since it is of interest to find the point where people responded "DIFFERENT" for instance 50 or 75% of the time, the psychometric function was calculated from the data using two methods: the classical procedure and the Signal Detection Theory procedure.

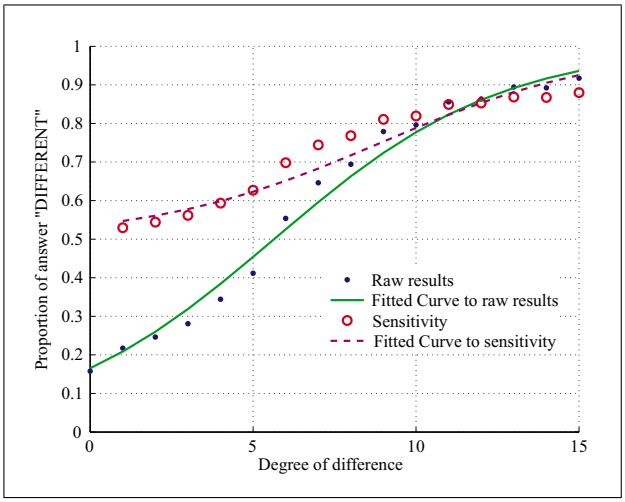


Figure 10. Raw results and fitted curve q_r of the psychoacoustical test to determine the JND.

5.2.1. Classical procedure

The classical procedure has been described by Berkson [31]:

First, a "logistic transformation" was calculated for each tone pair,

$$l_i = \ln\left(\frac{1 - q_i}{q_i}\right). \tag{6}$$

Then, a linear regression l_r was performed on the l_i vs i data,

$$l_r = mi + b, \tag{7}$$

obtaining $m = -0.287$ and $b = 1.6199$. The 50% point corresponding to the JND is calculated as

$$\text{JND} = \left|\frac{b}{m}\right|, \tag{8}$$

where m is the slope and b is the y-intercept of the regression line. This JND value corresponds to the point where the regression line crosses zero.

This gives a number between 0 and 15, which corresponds to the difference between the pair at which the subject answered 50% the same and 50% different. For the purpose of comparing the raw results with the fitted curve, the regression line was "detransformed" back to proportions q_r using

$$q_r = \frac{1}{1 + e^{l_r}}. \tag{9}$$

Figure 10 shows the raw results (represented as dots) and the detransformed fitted curve q_r (represented by the solid curve), which corresponds to the psychometric function. It can be seen from the Figure that the 50% JND is 5.6. Likewise, the 75% JND is 9.5.

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

5.2.2. Signal Detection Theory procedure

It is worth noting that the proportion of "DIFFERENT" answers was never zero, even when the control pair was presented. This situation is what in Signal Detection Theory (SDT) is called a "false alarm" [32]. It is the result of different listener's strategies. Some listeners will tend to answer "DIFFERENT" when they are in fact guessing (liberal criterion), while others will tend not answer "DIFFERENT" unless absolutely sure (conservative criterion). There are in fact 4 possible combinations of responses:

- The presented pair was the same and the listener responded "THE SAME": correct rejection
- The presented pair was the same and the listener responded "DIFFERENT": false alarm
- The presented pair was different and the listener responded "THE SAME": miss
- The presented pair was different and the listener responded "DIFFERENT": hit

One way of taking into account the listener's strategy, is by calculating what is called "sensitivity" or proportion of correct answers $P(C)$ [27]:

$$P(C) = P(h) \cdot P(s) + \big(1 - P(fa)\big) \cdot P(d), \qquad (10)$$

where $P(h)$ is the proportion of hits, $P(s)$ is the proportion of times when the presented pair was the same, $P(d)$ is the proportion of times when the presented pair was different and $P(fa)$ is the proportion of false alarms. For each difference degree, the sensitivity is calculated. If $P(s) = P(d)$ then equation 10 is reduced to the following:

$$P(C) = \frac{P(h) + \big(1 - P(fa)\big)}{2}. \qquad (11)$$

The data obtained from this psychoacoustical test shows that $P(fa)$ =0.1577, which is the proportion of times when the listeners answered "DIFFERENT" when presented with the control pair. For each degree of difference, the sensitivity was calculated using equation 11. The sensitivity results in values which vary between 50% and 100%, and increase as the difference degree between the signals of the presented pair increases [27]. These sensitivity values are shown in Figure 10 as circles. A curve can be fitted to this data in much the same way as described before, but first having scaled the sensitivity values as

$$P(C)_{\text{scaled}} = \big(P(C) - 0.5\big) \cdot 2. \qquad (12)$$

The fitted curve (detransformed and descaled) is shown in Figure 10 as broken curve. The midpoint of this curve corresponds then to the 75% JND, and is calculated to be 8.9, in contrast with the previous 75% JND value obtained with the classical procedure, which was 9.5.

The 75% JND obtained by the SDT method was divided by 30, so as to give a percentage of the amount of difference that the sounds had in comparison to what is shown in Figure 4. The JND was calculated to be 0.30, which corresponds to a pair of sounds whose harmonic differences correspond to approximately 30% of what is shown in Figure 4.

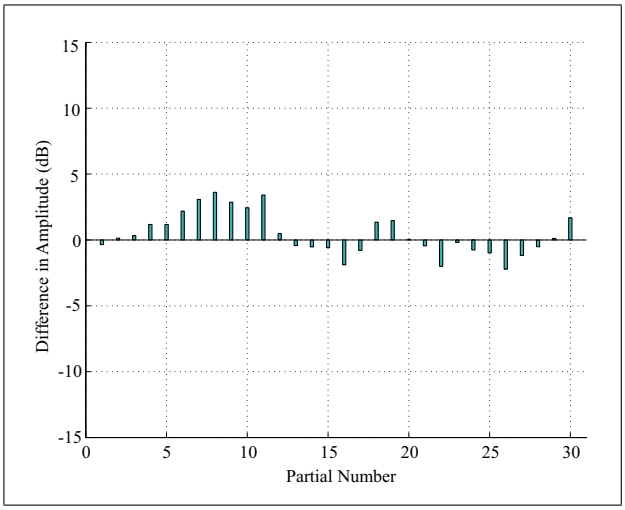


Figure 11. Spectral differences in harmonic content between the reference sound and the sound that corresponds to the 75% JND obtained in this experiment.

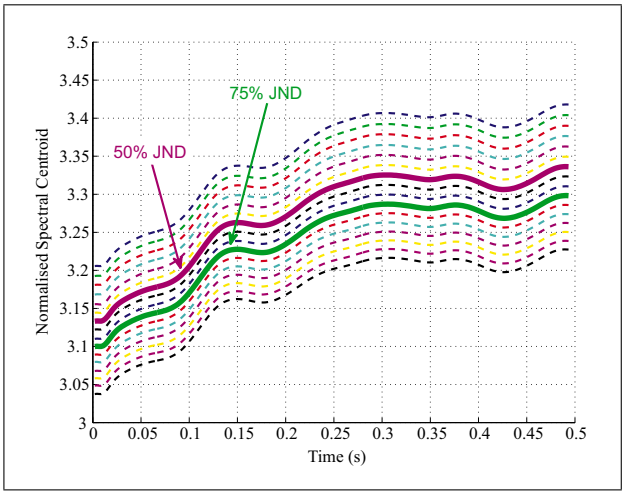


Figure 12. Spectral centroid variation over time of signals used for the psychoacoustical test. Solid lines correspond to the 50% and 75% JND's.

Figure 11 shows the differences in spectra between the reference sound and the sound corresponding to the 75% JND. The mean of the absolute difference shown in this figure was calculated to be 1.28 dB. Similarly, the 50% JND is found to be 0.86 dB

5.3. Just noticeable difference in spectral centroid

The spectral centroid variation over time of all 16 signals used in the previous psychoacoustical test is shown in Figure 12. The top curve corresponds to the reference signal. As each subsequent signal is modified slightly, the spectral centroid goes down gradually from one signal to the next.

The 50% and 75% JND found in the previous section were 5.6 and 8.9 respectively. To find which curve corresponds to the signal which have 50% JND, we count 6 curves below the reference (or 9 for the 75% JND). These two curves are marked with a solid thicker line in the Figure.

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

The mean of the absolute difference in spectral centroid between these signals and the reference were calculated to be 0.08 for the 50% JND, and 0.11 for the 75% JND. These values are in agreement with the results reported by Kendall and Carterette [13] (0.11 for musicians and 0.15 for non musicians). This is an interesting result, as Kendall and Carterette used artificial signals for their experiments: the signals they used as standards consisted of a fundamental of 440 Hz, and 4 upper harmonics, and their centroids were 2.2, 3 and 3.8. For each of these, they constructed signals that differed in centroid from the standards in steps of 0.02. In this paper the signals were taken from real recordings of trombone sounds, keeping all harmonics and high frequency components. The fact that the experiment shown in this paper agrees with their results provides evidence to support the assumption that the JND they reported are general and that they can be applied to other harmonic sounds.

6. Conclusions

Can the difference in sound of an instrument played with two different settings be perceived? In an attempt at answering this question, the objective of this paper was to find the JND in spectral content between two sounds that came from having recorded a trombone played with two different mouthpieces. In order to do this, a synthesis method was developed, whereby signals taken from trombone recordings were resynthesised. A 2AFC psychoacoustical test was done to make sure that the synthesised signal was indistinguishable from the original signal. With this synthesis method, the following attributes of the original sound were included in the synthesised signal: the relative amplitudes of the harmonics of the signal, the high frequency aperiodic components, and the slow frequency and amplitude modulation induced by the player.

Using this synthesis method, a signal from a recording of a trombone played with two different mouthpieces was synthesised, its harmonic amplitudes manipulated through spectral morphing, and arranged in pairs with varying degrees of difference in spectral content. A psychoacoustical test where pairs of sounds were presented in randomised order was performed in order to find the JND: the mean of the absolute difference in amplitude of the harmonics was found to be 0.86 dB for the 50% JND, or 1.28 dB for the 75% JND. Likewise, the JND of the mean of the absolute difference in normalised spectral centroid was also found: 0.08 for the 50% JND, and 0.11 for the 75% JND.

These results are in agreement with the results reported by Kendall and Carterette [13] (0.11 for musicians and 0.15 for non-musicians). This provides evidence to support the assumption that the JND in spectral centroid that they found in their psychoacoustical test using artificial signals with different degrees of spectral centroid are also valid for signals taken from the produced sounds of real musical instruments.

The results found in this paper only apply to the trombone signals studied, and it would be problematic to extrapolate that to other musical instruments. In order to do that, the procedure described here can be done with other instruments and compare whether those results agree with what was found here and by Kendall and Carterette [13]. Also, there are other parameters that have been correlated to the perception of timbre that were not considered here, such as the spectral flux, spectral irregularity and spectral shape [19].

Acknowledgement

I would like to present my deepest gratitude to Dr. Joël Gilbert for his willingness to play the trombone for the purpose of this study.

I would also like to thank Dr. Stephen Malloch for his suggestion about the method for loudness level calculation, Prof. James Beauchamp for his useful comments about the use of his program SNDAN, and Dr. Maarten van Walstijn for the fruitful discussions that led to the idea of the synthesis method described in section 4.2.

This work was supported by the CONACYT grant number 146702.

References

[1] J. M. Grey: Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America **61** (1977) 1270–1277.

[2] R. A. Kendall, E. C. Carterette, J. M. Hajda: Perceptual and acoustical features of natural and synthetic orchestral instrument tones. Music Perception **16** (1999) 327–364.

[3] J. R. Miller, E. C. Carterette: Perceptual space for musical structures. Journal of the Acoustical Society of America **58** (1975) 711–720,.

[4] K. W. Berger: Some factors in the recognition of timbre. Journal of the Acoustical Society of America **36** (1964) 1888–1891.

[5] E. L. Saldanha: Timbre cues and the identification of musical instrument. Journal of the Acoustical Society of America **36** (1964) 2021–2026.

[6] P. Iverson, C. L. Krumhansl: Isolating the dynamic attributes of musical timbre. Journal of the Acoustical Society of America **94** (1993) 2595–2603.

[7] J. M. Grey, J. W. Gordon: Perceptual effects of spectral modifications on musical timbres. Journal of the Acoustical Society of America **63** (1978) 1493–1500.

[8] J. M. Grey, J. A. Moorer: Perceptual evaluations of synthesized musical instrument tones. Journal of the Acoustical Society of America **62** (1977) 454–462.

[9] A. Horner, J. Beauchamp, R. So: Detection of random alterations to time-varying musical instrument spectra. Journal of the Acoustical Society of America **116** (2004) 1801–1810.

[10] S. McAdams, J. W. Beauchamp, S. Meneguzzi: Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. Journal of the Acoustical Society of America **105** (1999) 882–897.

[11] A. Nishimura, M. Kato, Y. Ando: Effect of the fluctuations of harmonics on the subjective quality of flute tone. – In: Proceedings of the Forum Acusticum Sevilla 2002. J. S. S. A. Calvo-Manzano, A. P. López (eds.). Sevilla, Spain, 2002, Sociedad Espa nola de Acústica.

[12] J. W. Coltman: Just noticeable differences in timbre of the flute. Catgut Acoustical Society Journal **3** (1996) 26–33.

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only

[13] R. A. Kendall, E. C. Carterette: Difference threshold for timbre related to spectral centroid. Proceedings of the 4^{th} International Conference on Music Perception and Cognition, Montreal, Canada, 1996, 97–101.

[14] H. A. K. Wright, D. M. Campbell: The influence of the mouthpiece on the timbre of cup-mouthpiece wind instruments. Proceedings of the International Symposium on Musical Acoustics, Leavenworth, WA, USA, 1998, 159–164.

[15] M. Campbell, C. Greated: The musician's guide to acoustics. Schimer Books, 1987.

[16] A. H. Benade: Fundamentals of musical acoustics. Oxford Univesity Press, 1976.

[17] B. R. Glasberg, B. C. J. Moore: A model of loudness applicable to time-varying sounds. Journal of the Audio Engineering Society **50** (2002) 331–342.

[18] W. Jesteadt, C. C. Wier, D. M. Green: Intensity discrimination as a function of frequency and sensation level. Journal of the Acoustical Society of America **61** (1977) 169–177.

[19] J. M. Hajda: The effect of dynamic acoustical features on musical timbre. – In: Analysis, synthesis, and perception of musical sounds. J. W. Beauchamp (ed.). Springer Science+Business Media, LLC, 2007, 250–268.

[20] J.-C. Risset, D. L. Wessel: Exploration of timbre by analysis and synthesis. – In: The psychology of music. D. Deutsch (ed.). Academic Press, 1982, 25–58.

[21] J. W. Beauchamp: Unix workstation software for analysis, graphics, modification, and synthesis of musical sounds. Audio Engineering Society, 1993, Preprint 3479.

[22] J. W. Beauchamp: Analysis and synthesis of musical instrument sounds. – In: Analysis, synthesis and perception of musical sounds: The sound of music. J. W. Beauchamp (ed.). Springer, 2007, 1–89.

[23] E. Tellman, L. Haken, B. Holloway: Timbre morphing of sounds with unequal numbers of features. Journal of the Audio Engineering Society **43** (1995) 678–689.

[24] L. Haken, K. Fitz, P. Christensen: Beyond traditional sampling synthesis: Real-time timbre morphing using additive synthesis. – In: Analysis, synthesis and perception of musical sounds. J. W. Beauchamp (ed.). Springer Science+Businees Media, LLC, 2007, 122–144.

[25] A. B. Horner, J. W. Beauchamp, R. H. Y. So: Detection of time-varying harmonic amplitude alterations due to spectral interpolations between musical instrument tones. Journal of the Acoustical Society of America **125** (2009) 492–502.

[26] B. C. J. Moore: An introduction to the psychology of hearing. 4th edition. Academic Press, 2001.

[27] W. A. Yost: Fundamentals of hearing: An introduction. Academic Press, 200.

[28] A. E. Brown: Measurement of auditory thresholds. Journal of the Acoustical Society of America **38** (1965) 86–92.

[29] H. Levitt: Transformed up-down methods in Psychophysics. Journal of the Acoustical Society of America **49** (1970) 467–477.

[30] D. Deutsch: The psychology of music. Second edition. Academic Press, 1999.

[31] J. Berkson: Application of the logistic function to bioassay. Journal of the American Statistical Association **39** (1944) 357–365.

[32] D. M. Green: Application of detection theory in psychophysics. Proceedings of the IEEE **58** (1970) 713–723.

Document downloaded by @DAEL, 2026-11-25 08:02:51 - personal use only