This item is the archived peer-reviewed author-version of:

Average middle ear frequency response curves with preservation of curve morphology characteristics

Reference:

Gladiné Kilian, Dirckx Joris.- Average middle ear frequency response curves with preservation of curve morphology characteristics
Hearing research - ISSN 0378-5955 - 363(2018), p. 39-48
Full text (Publisher's DOI): <https://doi.org/10.1016/J.HEARES.2018.02.005>
To cite this reference: <https://hdl.handle.net/10067/1495270151162165141>

Average middle ear frequency response curves with preservation of curve morphology characteristics

Gladiné Kilian^{1, a)}, Dirckx Joris J.J.^{1, b)}

¹University of Antwerp, Laboratory of Biophysics and Biomedical Physics, Groenenborgerlaan 171, 2020 Antwerp, Belgium

^{a)} Corresponding author: Kilian.Gladine@uantwerpen.be

^{b)} Joris.Dirckx@uantwerpen.be

Abstract

For the validation of modelling results or the comparison of middle ear interventions, such as prostheses placement, average responses of middle ear vibrations are needed. One such response is the amplitude and phase of the vibration of the stapes footplate as a function of frequency. Average responses and their standard deviation are commonly obtained by calculating the mean of a number of measured responses at each frequency. A typical middle ear magnitude response curve shows a number of distinct peaks, and the location of these peaks varies between ears. By simply taking an average along the magnitude or phase response axis, the typical fine structure of the response curve is flattened out, delivering an average curve which no longer has the typical morphology of an individual response curve. This paper introduces methods to avoid this problem by first aligning the typical curve features along the frequency axis prior to calculating the average along the magnitude or phase axis, resulting in average magnitude and phase curves which maintain the typical morphology of the curve obtained for an individual ear. In the method, landmark points on the response magnitude curves are defined and the frequencies at which these points occur are averaged. Next, these average frequencies are used to align the landmark points between curves, prior to averaging values along the magnitude or phase axes. Methods for semi-automatic and manual assignment of landmark points and curve alignment are presented. After alignment, the correspondence between the original landmark frequencies and aligned frequencies is obtained together with the warping function which maps each original magnitude curve to its aligned version. The phase curves are aligned using the warping functions determined from the corresponding magnitude curves. Finally, a method is proposed to compare the data set of an individual

Abbreviations:

MLA: Manual landmark assignment

SLA: Semi-automatic landmark assignment

measurement or model result to an aligned average curve in terms of magnitude and frequency by applying the alignment procedure to the individual curve.

1 Introduction

Frequency response curves of ossicular vibrations provide important data for validation of middle ear modelling results. As many models are built using average material parameters and shapes, it is common to compare model outcome to average response curves (De Greef et al., 2017). The most common way to obtain such a response curve is to average magnitude and phase values for each measured (or interpolated) frequency. Such averages give a good representation of the general frequency response of the middle ear (e.g. Rosowski et al., 2007).

Each individual response curve has a specific morphology, with local maxima that can be attributed to some extent to specific physiologic parameters or characteristic properties of the system, such as the resonance frequency of the ear canal (Keefe et al., 1993; Stinson, 1990) or the first ossicular mode (Homma et al., 2009). As shapes, masses and other parameters of middle ear structures vary from one individual to the next, both the height and frequency location of local maxima differ. By averaging along the magnitude axis (or along the phase axis), the sharpness of these local maxima diminishes because for each individual specimen the frequency location of the maximum is slightly different. As a result, the average response curve taken over many specimens shows a much less detailed structure than an individual response curve, and the typically peaked morphology of the curve tends to flatten out towards a broadly spread maximum at a center frequency. These less-pronounced peaks/troughs could easily be misinterpreted by modelers to mean that the middle-ear is more damped than it really is.

Therefore, it would be useful to generate average response curves which preserve the typical morphology of the individual response curve. In this paper we present a procedure to first align the typical morphologic features of response curves along the frequency axis prior to taking the cross sectional average along the magnitude or phase axis.

2 Materials and Methods

2.1 Input data

To demonstrate the new processing method, we used stapes vibration data obtained from human temporal bones in previous work (Niklasson, 2017). In these experiments, the stapes footplate velocity was measured from the medial side at the center of the footplate after removal of

the inner ear, and sound stimulation was presented through the intact ear canal. The response curves therefore show the transfer function of the middle ear system without cochlear loading. Measurements were taken at 16 frequencies per octave in a frequency range of 0.125-8kHz. As the processing of the data benefits from using more densely sampled curves, the curves were resampled using smoothed spline approximation to obtain a set of 500 logarithmically spaced data points.

2.2 Assigning landmarks

The alignment algorithm makes use of landmark points which are to be defined on individual curves (Kneip et al., 1992). In this section, we describe two ways to assign such landmark points. These landmark points will be used to align the vibration magnitude and phase curves.

2.2.1 Manual landmark assignment (MLA)

The least complex way to assign landmarks to the response curves is by choosing them manually. This is done by visually inspecting the data curves, choosing a curve feature, and noting the frequency at which this feature occurs for each curve. This can be done for several features. If a feature is not present in an individual curve, the landmark point is not assigned.

Up to 5 landmark points were used, which will be described in detail in the results section.

2.2.2 Semi-automatic landmark assignment (SLA)

In semi-automatic assignment, peaks in the curve are automatically detected. Then, conditions are set to filter the found peaks such that only the peaks of interest remain.

To find the peaks we have used the MATLAB function *findpeaks*. We then applied mathematical conditions to limit the number of found peaks to three landmarks. The conditions for each landmark were:

- The peak frequency was required to be lower than 1.5 kHz. In addition the peak frequency was required to be close to 1 kHz and preferred to have a high peak prominence. These last two conditions were applied using weighting. This landmark is indicated by a circle in figure 2.
- The peak with highest amplitude (indicated by an asterisk in fig. 2).
- The peak frequency was required to be higher than that of the peak with highest amplitude. In addition, the peak frequency was required to be the closest to 5 kHz (indicated by a diamond in fig. 2).

The resulting landmarks are displayed in the results section in figure 2. And the corresponding frequencies are given in table 1.

2.3 Curve alignment

2.3.1 Warping functions

In this section, we describe different warping functions and their properties. Warping functions are used to transform the data along the frequency-axis. The original frequency-axis should be projected on a new frequency-axis such that the landmarks of all curves are aligned.

The warping function $X(f)$ needs to be monotonic because each frequency f needs to be projected onto a unique frequency and the projected frequencies need to remain monotonically increasing. In other words, aligning the curves could mean we either shift them, and/or stretch/squeeze them but we cannot change the order of curve segments. As an analogy, for time-related data this criterion would imply the preservation of causality.

Several types of functions obey this criterion. A first example is the linear warping function.

$$X(f) = \alpha + \beta f \quad (1)$$

If $\alpha > 0$ the curve will shift to lower frequencies. When $\beta > 0$ the curve will be squeezed together; conversely, when $\beta < 0$, the curve will be stretched out. This type of warping function does not fix the end points of the curves and thus warped frequencies can exceed the minimal and maximal value of the original frequency range. If this occurs, data has to be estimated for frequencies outside the measured frequency range. This is no problem if all curves converge to a constant value or a similar slope at the boundaries of the data set. For our data set, this is not the case. At the upper boundary of the measured frequency range, some curves go up, and some go down. Extrapolation of the data will result in very different values and curve shapes for different curves and thus the use of a linear warping function is not advisable.

A second class of warping functions, satisfying monotonicity, is

$$X(f) = \frac{(f_{MAX} - f_{MIN}) \int_{f_{MIN}}^f \exp(g(s)) ds}{\int_{f_{MIN}}^{f_{MAX}} \exp(g(s)) ds} + f_{MIN} \quad (2)$$

for which $g(s)$ is constructed using a finite-dimensional basis $g(s) = \sum_{i=0}^q \gamma_i w_i(s)$ with $w_i(s)$ a q -dimensional basis, γ_i the coefficients and s is a dummy variable which takes on the value of the frequencies determined by the integral boundaries, similar to James (2007). A more detailed explanation on the use of the finite-dimensional basis can be found in appendix A. This warping function has the properties of a cumulative distribution function, but it is scaled and shifted such that the boundary values of the frequency range are projected on themselves (i.e.

$X(f_{MIN}) = f_{MIN}$ and $X(f_{MAX}) = f_{MAX}$). When the boundary values of the data range are fixed, possible need for extrapolation of the data is ruled out. A disadvantage is that rigid translations are not possible, the curve will always be deformed unless all γ_i are zero, then the warping function reduces to $X(f) = f$ (i.e. an automorphism).

We used the type of warping function defined by equation 2 to achieve the results shown in this paper. The q -dimensional basis was chosen to be constructed by third order basis splines with a dimensionality q of 100. A higher dimensionality increases the warping function flexibility, allowing for local warping. Other types of warping functions which fix the frequency boundaries and have sufficient flexibility should return very similar results.

2.3.2 Alignment cost function

We consider the curves to be aligned when the landmarks of the curves are aligned. For each curve we need to find the warping function that achieves this. When using warping functions of the type shown in equation 2, the problem is reduced to determining the coefficients γ_i .

The determination of the γ_i 's is done by minimizing a cost function C similar to James (2007) consisting of a few components (Q, R, S).

$$C = \lambda_Q Q + \lambda_R R + \lambda_S S + \dots \quad (3)$$

The parameters λ_Q, λ_R and λ_S are weights put on the individual cost contributions. Q is defined by the deviation of each landmark frequency from the geometric mean frequency of that landmark calculated using all curves for which that landmark has been assigned.

$$Q = \sum_{j=1}^P |f_j^i - \bar{f}_j| \quad \text{with } \bar{f}_j = \left(\prod_{i=1}^{N^j} f_j^i \right)^{\frac{1}{N^j}} \quad (4)$$

In equation 4, the index i represents the i th curve from the total number of N^j curves for which landmark j was assigned and P is the number of landmarks assigned on curve i . Note that there is no summation over the number of curves in Q . The optimization procedure aligns one curve at a time, thus Q only needs to contain landmark frequencies of one curve at a time. The average frequency was determined using the geometric mean since the frequency distribution of the landmark features more naturally fits a logarithmic distribution.

R is the average deviation of the warped frequencies $X(f_k)$ from their original frequencies f_k (eq. 5). This term will ensure that the warping functions will not go to extremes and maintain the

original shape of the curve as much as possible. It can also improve the convergence by forcing the algorithm to use smaller step sizes in the minimization.

$$R = \frac{1}{K} \sum_{k=1}^K |X(f_k) - f_k| \quad (5)$$

In equation 5, K is the total number of frequencies of each curve, not just the number of landmark frequencies. If the chosen warping function contains a shift term X_0 , equation 5 can be adapted to $R = \frac{1}{K} \sum_{k=1}^K |X(f_k) - f_k - X_0|$. This would force the algorithm to achieve an alignment using mainly the shift term and thus minimizing the amount of curve deformation.

Conditional statements can also be used as shown in equation 6. The term S is a hard condition to avoid too strong local warping. When there is no warping, the derivative of the warping function $X'(f)$ is 1 for all frequencies.

$$S = \begin{cases} +\infty & \text{if } |X'(f_k) - 1| > \text{tolerance} \\ 0 & \text{if } |X'(f_k) - 1| < \text{tolerance} \end{cases} \quad (6)$$

The weights $(\lambda_Q, \lambda_R, \lambda_S)$ can be set by experimentation and are dependent on the user. This dependency exists due to the goals of terms R and S . Which amount of (local) warping is *too much*? Additional cost terms can be used to increase control over the alignment procedure.

In this study, the convergence of the alignment of one curve using SLA landmarks was improved using $\lambda_R = 0.01$. The alignment of other curves did not pose any problems regarding an excessive amount of (local) warping. For this reason, λ_S was set to zero. Parameter λ_Q was set to one.

2.3.3 Comparing a single data curve with an aligned average curve

In this section, we describe how we can compare a single curve with aligned average data. The curve features of the single curve are not necessarily aligned with those of the aligned average data. Comparing the magnitude of both datasets at each frequency individually would therefore be an unfair comparison. We can use the alignment procedure to align the curve features of the single curve to that of the aligned average data. It is important to note that only the single curve will be warped; the aligned average data remains untouched. We can then compare both datasets in terms of magnitude by comparing the aligned curves but we can also compare them in terms of frequency by inspecting the warping functions.

To demonstrate this, we isolated curve number 5 from the six curves (see fig.1 for the numbering of the curves). Curve number 5 was chosen to be isolated as it was the most different in comparison with the other curves. A new aligned average and standard deviation were calculated using only the

five remaining curves. Next, a cross-sectional average and standard deviation of the five warping functions, originating from the alignment procedure of the five remaining curves, were calculated. Then, curve number 5 was aligned with the aligned average using landmarks determined by MLA. After this, the aligned curve number 5 and the aligned average were compared. The difference along the magnitude-axis was inspected using the aligned curves themselves, while the difference along the frequency-axis was inspected by comparing their warping functions.

3 Results

3.1 Assigned landmarks

The manually assigned landmarks are shown in figure 1. The first landmark (indicated by a circle in fig.1) was assigned to the local maximum after the initial increasing slope, ignoring local maxima below 600 Hz. The second landmark (square in fig. 1) was the next local maximum between 1-2 kHz. This landmark was not clearly identifiable for curve number 6. The third landmark (asterisk in fig. 1) was assigned to the global maximum. The fourth landmark (diamond in fig. 1) was assigned to the first prominent local maximum after the downward slope after the global maximum. The fifth and last landmark (triangle in fig. 1) was assigned to the local maximum at frequencies higher than the fourth landmark. For curve 4 and 6, this local maximum was not found. To avoid confusion, the landmarks will be referred to by their symbol only from this point on.

Figure 1: Landmarks for each frequency response magnitude curve resulting from manual landmark assignment (MLA). For specimens 4 and 6, not all landmarks could be assigned.

The semi-automatic assignment of three landmarks resulted in similar landmarks to those of manual assignment (fig. 2). The landmark points of the SLA method which correspond to the landmark points of the MLA method are indicated using the same symbol. Only 2 significant differences are present, the circle landmark is placed at a higher frequency for curve 4 when using SLA as opposed to MLA, while the diamond landmark ($\diamond$) is assigned to a lower frequency. The numerical values of the landmark frequencies for both methods are given in table 1.

Figure 2: Landmarks for each frequency response magnitude curve resulting from semi-automatic landmark assignment (SLA). For corresponding landmarks the same marker-types were used as for the manually assigned landmarks (fig. 1).

Method	Landmark Curve	o	□	*	◇	Δ
MLA	1	988	1698	4211	5452	6441
	2	901	1523	3810	4732	6076
	3	750	1216	3335	4732	7001
	4	1013	1390	3419	6550	-
	5	843	1268	2685	4390	5976
	6	1091	-	4577	6335	-
SLA	1	979	-	4211	5498	-
	2	901	-	3810	4654	-
	3	756	-	3335	4693	-
	4	1378	-	3419	5544	-
	5	857	-	2685	4390	-
	6	1091	-	4577	6335	-

Table 1: Frequencies in Hertz of the landmarks assigned using MLA and SLA. Landmarks that were not assigned are depicted as a dash (-). Landmarks that have been placed at a significantly different frequency for the MLA and SLA methods are shaded.

3.2 Curve alignment

Figure 3a shows the original magnitude response curves for all 6 specimens. The average response curve (blue line) is obtained by just averaging the magnitude values. Figure 3b shows the response curves after alignment of the landmark points determined by MLA. The aligned MLA average is shown in red. Figure 3c shows the result using SLA landmarks, the aligned SLA average is shown in green. The traditional cross-sectional, MLA aligned and SLA aligned averages are shown again in figure 3d for comparison. After alignment of the landmarks assigned by either method, the average curve shows more distinct features which are also seen in the individual curves (fig. 3). The rising slope below 0.7 kHz is similar to that of the traditional cross-sectional average curve. A small difference is present at 0.6 kHz, the original average curve makes a slightly more noticeable dip. The first local maximum is less flat than that of the original average curve. This is also true for the local maximum in the 1-2 kHz region. The largest difference is seen at the global maximum. The MLA and SLA aligned averages show a distinctive peak for which the amplitude is higher than for the original average. For frequencies beyond the maximum response, the MLA and SLA aligned averages show two distinct peaks instead of one peak and a shoulder. All averages converge to the same values at the boundaries of the measured frequency range due to the boundary conditions of the warping function (section 2.3.1).

Figure 3: Individual and cross-sectional average frequency response magnitude of (a) original data, (b) aligned data using MLA, (c) aligned data using SLA. (d) Comparison of cross-sectional averages of (a), (b) and (c).

The difference between the MLA and SLA averages are small. There is a small discrepancy in the locations of the first and second local maximum of the SLA and MLA aligned averages (56 Hz and 73 Hz respectively) and a small difference in amplitude of the last two local maxima (-0.8dB and -0.6dB respectively).

The warping functions associated with the MLA and SLA average are shown in figure 4. The MLA warping functions are a little more straight at the lower frequencies, but some of them are broader in the 4-6 kHz range.

Figure 4: Warping functions when using MLA (a) and using SLA (b).

The effect of applying the warping functions to the phase curves (fig. 5) is less pronounced than for the magnitude curves. The most obvious change occurs again at around 4 kHz. The aligned averages show a steep phase jump. Between 1 and 2 kHz, the MLA and SLA aligned averages show small nuances indicating the local maxima. This is not the case for the original average for which the phase curve is almost a straight line in the 1-2 kHz range.

Figure 5: Individual and cross-sectional average frequency phase response of (a) original data, (b) aligned data using

manual landmark assignment (MLA), (c) aligned data using semi-automatic landmark assignment (SLA). (d) Comparison of cross-sectional averages of (a), (b) and (c).

4 Discussion

4.1 Landmark assignment

The manual assignment of landmarks is robust and grants a lot of freedom to the user. It is relatively simple to visually recognize and mark features of the curves. The overall shape of each curve and the fact that all curves represent the same variable is taken into account. It can always be decided that a landmark, associated with a certain curve feature, will be assigned to a curve or not. If the associated curve feature is not present, the landmark is simply not assigned for that curve.

The semi-automatic assignment of landmarks offers speed as its main advantage. The biggest drawbacks are the lack of flexibility and that landmarks can easily be placed at locations which do not correspond with intuition. This was the case for two landmarks (O, $\diamond$) of curve 4 (fig. 2). The circle landmark of curve 4 was assigned to the resonance peak between 1-2 kHz, which is actually the square landmark between 1-2 kHz for the manual landmark assignment. The cause is the prominence criterion set in the SLA algorithm. The prominence of the resonance between 1-2 kHz was higher than that of the resonance at 1 kHz; giving the 1-2 kHz resonance a higher weight in the SLA algorithm and thus making this resonance more likely to be chosen as the requested landmark. However, if prominence was not taken into account, the same would have happened for curve 3 instead, for which the resonance between 1-2 kHz is the closest to 1 kHz. This shows that it is hard to design SLA criteria that result in the same landmarks as the MLA method for all curves. Overall, 16 out of 18 semi-automatic landmarks using SLA closely match with the corresponding landmarks using MLA. The SLA method performance is also less robust when looking for the two peaks at around 5 kHz. Additionally, it is very complicated to design criteria such that the SLA method can decide if a landmark is present or not. For this reason, the number of landmarks to be assigned using SLA was reduced from five to three.

More complicated methods to find corresponding data points that can be used as landmark points exist such as the minimum description length (Thodberg, 2003). We also tested this method but for our application, even the minimum description length method with the addition of curvature cost (Thodberg et al., 2003) did not seem to improve the results.

One could argue that the MLA method is subjective and different people could obtain different results. In SLA, different algorithms can also result in different landmarks. The presented SLA method inherently includes subjectivity by utilizing user defined criteria and criteria weightings. A fully automatic solution could be possible using machine learning. The use of machine learning could reduce or even completely remove the subjectivity. A large number of input data will be needed to train the decision algorithm. We believe that if a group of experts can agree on the chosen landmarks, and alignment result, then it does not matter if the landmarks were chosen manually or (semi-)automatically. To achieve this, it is advised not to assign many more landmarks than the amount that seems necessary. Preferably, the features chosen to determine the landmarks are features which are known to be existent for every curve. For our magnitude curves, this is the resonance peak at around 1 kHz, associated with a hinging ossicular motion (Homma et al., 2009) and the resonance of the ear canal at around 3-4 kHz (Keefe et al., 1993; Stinson, 1990). In contrary, resonances below 0.7 kHz are uncommon and not generally reported in literature. The most robust landmarks will be based on underlying knowledge of the physiology which leads to the existence of the specific feature in the curve.

Landmarks close to the frequency range boundaries (e.g. Δ , fig. 1) may not be ideal. For curve 4 and 6, the triangle landmark could perhaps be found just outside the measured frequency range. This reduces the validity of the averaged final local maximum. Four out of six data points are now averaged with two non-corresponding data points, which is similar to what happens with traditional cross-sectional averaging. The same applies for the data points at the fixed boundary frequencies.

If the data contains noise, choosing landmark features becomes more difficult. Using the MLA method, the most obvious landmark features (O, $\square$, *) can still be easily determined. Distinguishing between the triangle and diamond landmark would be harder, even impossible. If these two landmarks would appear as one broad peak in most curves due to noise, the center of this broad peak can be used as a landmark feature. The SLA method will perform worse when noise spikes are present. By smoothing the original data, the noise spikes are reduced and performance can be increased.

It is important to note that other landmark features than local maxima can be used. Local minima in the data have also been linked to physical phenomena within the middle-ear or experimental set-up (e.g. resonance at around 2 kHz in chinchilla middle-ear due to hole drilled in bulla for experimental measurements, Rosowski et al. 2006). Such curve features can also be used for the alignment process.

4.2 Curve alignment

Comparison between the traditional average and the MLA and SLA aligned averages

Both landmark assignment methods show good results after alignment (fig. 3). For the MLA and SLA aligned averages, a clear distinction can be made between the first and second resonance in the 0.9-2 kHz range. In contrary, the traditional cross-sectional average has a nearly constant value in that range due to the fact that both resonant peaks occurs at different frequencies for different curves.

Similarly, the traditional average showed a broad peak at the global maximum and just one local maximum and a shoulder at higher frequencies. This is due to the misalignment of the global maximum and overlap with the local maxima at higher frequencies. After alignment, this broad peak is resolved into one peak and two local maxima at higher frequencies, improving representation of the average curve shape and estimation of the average peak height. It should be noted that due to the frequency warping of the individual transfer functions the width of the curve peaks has been deformed and the aligned average curve does not show the mean width of the individual curve peaks and thus does not reflect the mean damping behavior.

Effect of dimensionality q

The dimensionality q of the basis used to construct the warping functions determines the flexibility of the warping functions. When q is large, the flexibility increases and local warping can be done. This can be observed in the warping functions, if they show multiple bending points, the flexibility is high (fig. 4). When q is low (e.g. $q < 50$), finer details such as the last two resonances are harder to align. When q is high (e.g. $q > 200$), the optimization takes longer and may result in undesired strong local warping. Although, this can be limited by the cost term S , which limits the amount of local warping. From experience, choosing $q \approx 100$ should suffice. The effect of choosing different values for q is shown in appendix B.

Effect of fixed frequency boundaries

Fixing the frequency boundaries is necessary when no accurate extrapolation of the data can be done at one or both boundaries. In our case, the data at the 8 kHz boundary showed different slope directions for different curves, making extrapolation unsuitable. However, if the data is predictable (e.g. convergence to a constant value or slope) and similar for all curves, it can be considered to include a shift term in the warping functions. A shift term can produce a reasonable alignment without deforming the curve, resulting in an improved preservation of the original shape after full alignment.

Aligning phase curves

Aligning the phase curves using the warping functions associated with the aligned magnitude curves has two advantages. First, the average has improved with respect to the traditional cross-

sectional average (fig. 5). Although the difference is less drastic than for the magnitude curves, some details of the individual curve shapes are now preserved. The individual phase changes corresponding to local magnitude maxima between 1-2 kHz and around 5 kHz are a good example of this. Second, the correspondence between the phase and magnitude data points is restored. For example, the steep phase jump around 4 kHz occurs at the same frequency as the global maximum of the magnitude curve.

Other alignment methods

Other alignment methods such as Procrustes alignment (Ramsay et al., 1998) and alignment by moments (James, 2007) were considered. However, the Procrustes method tries to align the curves by minimizing the square distance of the magnitudes to those of a target function. This target function is usually chosen to be the cross-sectional average of the curves. The focus on absolute magnitudes instead of location of curve features results in large deformations of the magnitude curves, especially at the rising slope below 1 kHz.

The alignment by moments makes use of curve feature functions which are integrated over the entire frequency range to produce moments. These moments are then matched by a minimization procedure, similar to that presented earlier (section 2.3.2.). Moments based on feature functions considering local maxima are mainly determined by the zero-crossings of the first and second derivative of the data. As a consequence, these moments are heavily dependent on the number of data points. Additionally, the moments are also dependent on the number of local maxima, which is not the same for all magnitude curves.

Additional remarks

The main purpose of the present paper was to present the advantages of pre-aligned averaging over traditional cross-sectional averaging. We based our average on six measured curves as these are the data we have currently available. To generate a truly representative average, much more response curves will need to be gathered. In collaboration with other authors we also hope to gather data of other response curves such as umbo vibration amplitude or stapes vibration amplitude in the presence of intact cochlear load, so that our method can be used to generate a set of standard reference curves. It is important to note that in order to generate a standard reference curve, the data sets which are used should all be obtained under the same experimental conditions. Specimens should for instance be prepared in the same way so that possible artifacts (e.g. an anti-resonance caused by making an access hole to the middle ear cavity) are taken into account in the same way between combined data sets.

4.3 Comparing data to aligned average data

For the purpose of demonstration we compared one measured curve to the frequency-aligned average of the other measured curves, and found that it significantly deviates from the average, possibly indicating it should be treated as an outlier. The same method of comparison can be used for validation of modelling results.

Figure 6a shows the magnitude of curve 5 in both its original and aligned state compared with the aligned magnitude average and standard deviation band calculated using all curves excluding number 5. For frequencies up to 1.5 kHz, curve number 5 does not lie within the standard deviation band of the aligned average. The same holds for frequencies above 6 kHz and the magnitude at the global maximum. We can conclude that for frequency ranges 0.125-1.5 kHz, 3.9 kHz and 6-8 kHz the magnitude is significantly below average.

Figure 6: (a) Comparison of magnitude curve 5 before and after alignment with the average aligned curve μ_{MLA} (remaining 5 curves used) and its standard deviation band σ_{MLA} . (b) Comparison of the warping function of curve 5 with the average warping function $\mu_{X(f)}$ and its standard deviation band $\sigma_{X(f)}$ of the 5 curves used to obtain the aligned average magnitude curve.

Similarly, the warping function used to align curve number 5 with the aligned average does not lie within the standard deviation band of the average of the warping functions, originating from the alignment procedure of the 5 curves, for a frequency range of 1-7.5 kHz, almost spanning the complete data range (fig. 6b). It can be concluded that the warping function deviates far from an

automorphism for a frequency range of 1-8 kHz. These observations possibly indicate that curve 5 should be treated as an outlier.

Finally, it needs to be noted that using standard deviation bands on the warping functions is not strictly correct and should be used with care. For example, the standard deviation band can violate the requirement of monotonicity of the warping functions. Still, it can be used as an indication of the agreement of frequency locations of curve features between the aligned average and an individual curve as shown in this example.

5 Conclusion

In this paper, we have introduced a method for calculating average middle ear frequency response magnitude and phase curves with preservation of curve characteristics. We have demonstrated this method using stapes footplate velocity data obtained from human temporal bones with intact ear canal and removed cochlea. We have proposed a manual and semi-automatic method of generating landmark points on individual magnitude curves. A warping function was chosen and employed to align individual magnitude curves. A minimization algorithm was implemented to optimize alignment of landmark points by frequency and additional terms to control the amount of warping were suggested. It was shown that the phase curves can be aligned accordingly using the warping functions determined from the corresponding aligned magnitude curves. We introduced a procedure to test if a data set lies within an acceptable range of the aligned average data on the magnitude (or phase) axis as well as on the frequency axis: this method can be used to validate models or to determine outliers in an experimental data set.

In future work, we plan to collect more data to improve this middle ear frequency response average. We will also collect data originating from temporal bones with intact inner ear and no ear canal, which are the conditions used in many models, to construct a frequency response average representing those conditions as well. We invite other researches to bring together data on middle ear vibrations to use the new averaging method for the generation of a set of reference curves.

6 Acknowledgements

Special thanks to Toon Huysmans and William Keustermans for suggesting literature and introducing us to shape analysis methods. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

7 References

- De Greef, D., Pires, F., Dirckx, J.J. 2017. Effects of model definitions and parameter values in finite element modeling of human middle ear mechanics. *Hear Res* 344, 195-206.
- Homma, K., Du, Y., Shimizu, Y., Puria, S. 2009. Ossicular resonance modes of the human middle ear for bone and air conduction. *J Acoust Soc Am* 125, 968-79.
- James, G.M. 2007. Curve Alignment by Moments. *Ann Appl Stat* 1, 480-501.
- Keefe, D.H., Bulen, J.C., Arehart, K.H., Burns, E.M. 1993. Ear-canal impedance and reflection coefficient in human infants and adults. *J Acoust Soc Am* 94, 2617-38.
- Kneip, A., Gasser, T. 1992. Statistical Tools to Analyze Data Representing a Sample of Curves. *Ann Stat* 20, 1266-1305.
- Niklasson, A. 2017. An optimal partial ossicular prosthesis should connect both to the tympanic membrane and malleus. A temporal bone study using Laser Doppler Vibrometry. *Otol Neurotol* (Accepted).
- Ramsay, J.O., Li, X.C. 1998. Curve registration. *J Roy Stat Soc B* 60, 351-363.
- Rosowski, J.J., Ravicz, M.E., Songer, J.E. 2006. Structures that contribute to middle-ear admittance in chinchilla. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 192, 1287-311.
- Rosowski, J.J., Chien, W., Ravicz, M.E., Merchant, S.N. 2007. Testing a method for quantifying the output of implantable middle ear hearing devices. *Audiol Neurotol* 12, 265-76.
- Stinson, M.R. 1990. Revision of estimates of acoustic energy reflectance at the human eardrum. *J Acoust Soc Am* 88, 1773-8.
- Thodberg, H.H. 2003. Minimum description length shape and appearance models. *Lect Notes Comput Sc* 2732, 51-62.
- Thodberg, H.H., Olafsdottir, H. 2003. Adding curvature to minimum description length shape models. In: Bangham, R.H.a.A., (Ed.), *British Machine Vision Conference*. BMVA Press. pp. 26.1-26.10.

8 Appendix

A. Finite dimensional basis and warping functions

The warping function given in equation 2 makes use of function $g(s)$ which is constructed using a finite dimensional basis $w(s)$. In this paper we have chosen to use a b-spline functions for $g(s)$. Such a b-spline function is defined as a linear combination of control points γ_i and b-spline basis functions $w_i(s)$: $g(s) = \sum_{i=0}^q \gamma_i w_i(s)$, for this reason we call the control points also coefficients. We make use of b-spline basis functions of the third order ($k = 3$). A lower order results in a less smooth warping function. An example of a b-spline basis is shown in figure 7, for the sake of clarity $q = 5$ is used.

Figure 7: Example of a third order ($k = 3$) b-spline basis $w(s)$ with dimensionality $q = 5$ defined on the frequency range of the data sets using logarithmically spaced elements of the knot vector T .

The coefficients γ_i completely determine $g(s)$ and thus $X(f)$. For example, if all γ_i are chosen to be equal, $g(s)$ is a constant. As a consequence, the integral in equation 2 will be a linear function. Since $X(f)$ is normalized to fit the desired frequency range $[f_{MIN}, f_{MAX}]$, only one linear function holds, which is the automorphism $X(f) = f$. Figure 8 shows the function $g(s)$ and the corresponding $X(f)$ for various values of the coefficients γ for $q = 5$.

Figure 8: Example of the b-spline function $g(s)$ for different values of γ (left) using a third order b-spline basis ($k = 3$) and dimensionality $q = 5$ and the associated warping functions $X(f)$ (right).

The MATLAB code for determining b-spline basis is available at the MATLAB file exchange website. To construct a b-spline basis, the order k of the basis functions and a knot vector T is required. The basis functions are defined on this knot vector T , which in our case spans the required frequency range. For a dimensionality of q , T contains $q + k + 1$ elements, but there are only q unique elements. The first and last element, f_{MIN} and f_{MAX} respectively, are repeated k times. The q unique elements were logarithmically spaced.

B. Effect of dimensionality q

Figure 9 shows the effect of choosing different values of the dimensionality q on the aligned average. It is clear that a lower flexibility will result in bad alignment. For $q = 1$, the alignment result is mediocre for every landmark feature, resulting in an increased number of resonance peaks. The average converges as q gets larger. Figure 10 shows the effect of q by displaying the aligned individual curves.

Figure 9: The effect of the dimensionality of the warping functions on the aligned magnitude average.

Figure 10: The effect of the dimensionality of the warping functions on the individual aligned magnitude curves.