J Am Acad Audiol 25:541–548 (2014)

Allowing for Asymmetric Distributions When Comparing Auditory Processing Test Percentage Scores with Normative Data

DOI: 10.3766/jaaa.25.6.4

Dani Tomlin*†
Harvey Dillon*‡
Andrea S. Kelly§**

Abstract

Background: Raw percentage scores can be transformed to age-specific Z scores, despite the asymmetric distribution of normative data using a process that is applicable to any percentage (or proportion)-based result.

Purpose: Normative values are generated for the commonly used dichotic digit and frequency pattern behavioral tests of auditory processing.

Study Sample: A total of 180 normal-hearing children aged 7 yr 0 mo to 12 yr 2 mo took part in this study.

Research Design: A transformation and regression method is incorporated that allows for the asymmetric distribution of normative results and the development of the response across the 7–12-yr-age range.

Data Collection and Analysis: Percentage correct scores were determined for each ear in the dichotic digit and frequency pattern tests, delivered at 50 dB HL. The scores were arcsine transformed, then regressed against using an exponential equation, providing an age specific estimated mean score. The residual error of the regression was then used to estimate age specific variance.

Results and Conclusions: The ability to express results along an age continuum (while accounting for the asymmetric distribution and significant developmental influences) as a standard unit across all ages enables a simplified expression of performance ability on a task.

Key Words: Auditory processing disorders, Z score

Abbreviations: AD = Anderson Darling; APD = auditory processing disorder; DDT = dichotic digits test; FPT = frequency pattern test

INTRODUCTION

Evidence suggests that a significant number of school-aged children may have difficulties on auditory processing tasks, with associated learning difficulties (Sharma et al, 2009). Standard practice in identifying such children has been to determine whether a child's performance is above or below a cutoff score for a given age band. Recommended cutoff criteria for the normal range has been 2 or 3 SDs below the mean (American Speech-Language-Hearing Association Task Force on Central Auditory Processing Concensus Development, 1996; American Academy of Audiology, 2010). Limitations occur, however, when this approach is used with percentage or proportion-based data (Studebaker, 1985)

The dichotic digits test (DDT; Musiek, 1983) and frequency pattern test (FPT; Musiek, 1994; Noffsinger et al, 1994) are standard inclusions as tests of binaural

---

*HEARing Cooperative Research Centre, University of Melbourne Department of Audiology, Hearing and Speech Sciences, Parkville, Australia; †University of Melbourne Department of Audiology, Hearing and Speech Sciences, Parkville, Australia; ‡National Acoustic Laboratories, Macquarie Park, Australia; §University of Auckland, Auckland, New Zealand; and **Auckland District Health Board, Auckland, New Zealand

Dani Tomlin, PhD, Department of Audiology and Speech Pathology, University of Melbourne, 550 Swanston St, Parkville, Australia, 3010; E-mail: dtomlin@unimelb.edu.au

Funding support for this project was provided by the The HEARing Cooperative Research Centre, established under the Cooperative Research Centres program, an Australian Government Initiative.

integration and temporal pattern processing, respectively, in many auditory processing test batteries, (Bellis, 1996, 2003; Singer et al, 1998; Bellis and Ferre, 1999; Kelly, 2007; Dawes et al, 2008; Sharma et al, 2009; Iliadou and Bamiou, 2012), and both were frequently reported as used in a survey of clinicians (Emanuel, 2002). A point of difference between these two tasks and others commonly seen in an auditory processing disorder (APD) test battery is that the final score is expressed as a percentage of correct responses, rather than a unit score such as decibel or millisecond. Currently, the only published normative data available to clinicians are norms derived from studies of specific populations in the United States and New Zealand (Musiek et al, 1985; Bellis, SD; Singer et al, 1998; Kelly, 2007) (see Table 1).

The first difficulty with this form of result is that of basing calculations on the mean and SD of an asymmetric distribution of data. Percentage scores are distributed binomially, rather than normally, around the mean (Studebaker, 1985). This results in an asymmetric distribution. The maximum score possible is 100%, which may be less than 2 SDs above the mean; however, a low percentage score may be many SDs below the mean but still within the range occupied by 95% of the data. Furthermore, the interpretation of a difference in score between 85% and 95% and 40% and 50% is not the same, because the test-retest variance is much smaller in the first case than in the latter. A transformation process is required that normalizes the distribution of the data by keeping the variance of data around the mean constant and which thus produces a linear scale of change in SD units. The double arcsine (or angular) transform (Eq. 1) is recommended to deal with percentage, or proportionate results. Examples of the use of the arcsine transform in speech and hearing literature are reported by Studebaker (1985):

$$\mathrm{T} = \mathrm{sine}^{-1}(\sqrt{(\mathrm{P.N}/(\mathrm{N}+1))}) + \mathrm{sine}^{-1}(\sqrt{((\mathrm{P.N}+1)/(\mathrm{N}+1))}), \tag{1}$$

where T is the transformed score and P is the proportion of items correct, i.e., bounded by 0 and 1 and N is the number of presentations (items tested).

A second limitation of a simple comparison with a cutoff score is the expression of performance purely as pass or fail (Jerger and Musiek, 2002; Wilson et al, 2004; Dillon et al, 2012). The first difficulty with this dichotomy is that it does not give any indication of the *degree* of skill deficit. The second problem with a dichotomy is that two scores differing by only a tiny degree can result in opposite results to each other. An alternative consideration of the results is as a Z score. This is an expression of result in SDs of the population under consideration relative to the mean of that population, as shown in Equation 2, where X is the percentage score obtained. For example, a Z score of –3.5 indicates performance 3.5 SDs below the mean. A score of +1 is 1 SD above the mean. Therefore, when results are expressed in this manner, it states the degree of the difficulty with the task relative to the normative population used:

$$\mathrm{Z} = (\mathrm{X} - \mathrm{Mean})/\mathrm{SD}. \tag{2}$$

A third complication in comparing an individual's score with that of a relevant population occurs when the mean population scores vary markedly with age. As examples, both binaural integration and temporal-ordering

Table 1. Summary of Published Normative Values for the DDT and FPT

		Dichotic Digit Test (Mean ± SD)		Frequency Pattern Test (Mean ± SD)	
Study	Age (yr)	Left Ear	Right Ear	Right Ear	Left Ear
Kelly (2007)	7–8	85.3 ± 9.0	87.9 ± 8.4	71.0 ± 19.4	72.6 ± 18.9
Singer et al (1998)	7	73.83 ± 6.4	74.2 ± 6.0		
	8	88.70 ± 5.93	91.7 ± 4.9		
Bellis (1996)	7	55	70		
(Cutoff scores)	8	65	75	42	42
Musiek (test manual) (cutoff scores)	7–8	65	75	40	40
Kelly (2007)	9–10	91.24 ± 4.4	93.2 ± 4.3	87.27 ± 8.4	85.2 ± 8.4
Singer et al, (1998)	9	91.41 ± 4.7	92.53 ± 5.8		
	10	92.28 ± 3.6	93.17 ± 4.7		
Bellis (1996)	9	75	80	63	63
(Cutoff scores)	10	78	85	78	78
Musiek (test manual) (cutoff scores)	9–10	75	80	65	65
Kelly (2007)	11–12	94.3 ± 4.3	94.4 ± 4.4	92.7 ± 8.4	91.4 ± 8.4
Singer et al (1998)	11+	96.0 ± 3.2	96.0 ± 3.2		
Bellis (1996)	11+	88	90	78	78
(Cutoff scores)	12			80	80
Musiek (test manual)	11–12	88	90	75	75

Notes: Normative values are shown as mean values ±1 SD and are both expressed as percentage scores. The Bellis and Musiek data are shown only as the cutoff scores, which equal 2 SDs below the mean.

(pattern-processing) skills of auditory processing have been demonstrated to undergo significant development in normal children, with responses not reaching maturity until approximately age 12 yr (Musiek, 1994; Bellis, SD; Singer et al, 1998; Baran and Musiek, 1999; Kelly, 2007). The usual solution is to express the results as mean and SDs of the percentage scores in bands of 1 or 2 yr (see Table 1). There are two difficulties with this. First, the progression of means, SDs, and the consequential cutoff scores rarely shows the smooth variation with age group that is expected with development. This is a result of the random-sampling effects within the relatively small samples of children on whom the normative data are based, which will occur unless the normative data group comprises several hundred participants. Second, when the rate of development is sufficiently steep, which commonly occurs in the younger age groups, the use of a single cutoff value for an entire year of age, and especially a 2 yr age bracket, can easily give misleading results.

An alternative approach to the interpretation of percentage scores is outlined in this study, along with a set of combined Australian and New Zealand population normative values for these two commonly used APD tasks. A regression method is contained in the analysis and transformation process of the percentage scores outlined, which accounts for the child's age as a continuous variable.

The aim of this study is to provide the clinician with a new interpretation of normative data for two commonly used auditory processing tasks. Furthermore, expression of results along a continuum, while accounting for the asymmetric distribution and significant developmental influences, as a standard unit across all ages is possible. This trend enables a clear expression of degree of deficit present. This methodology is directly transferrable to any task generating percentage scores where the mean performance of the reference population varies with age.

PARTICIPANTS AND TEST MATERIALS

A total of 180 participants contributed to the normative data set, from two groups. The first group represents an Australian population, recruited through the general community and metropolitan primary schools, comprising 50 children (19 males and 31 females), between ages 7.0 and 12.2 yr. The mean group age was 9.1 ± 1.4 yr. The second group (n = 130) is from a New Zealand population. The data collection and characteristics for this group are outlined in an earlier publication (Kelly, 2007).

Inclusion criteria included no reported concerns regarding hearing or listening ability. Further criteria for inclusion were no reported parental concern with academic progress or learning ability; or a history of significant middle ear disease, as defined as three or more reported episodes of middle ear effusion and/or history of ventilation tube placement.

Peripheral hearing was assessed in quiet or soundproof conditions using standard audiometric procedures (either at the child's school, the home, or the University of Melbourne Audiology Clinic). Hearing thresholds were required to be 15 dB HL or better at octave frequencies from 500 Hz to 4 kHz. Evidence of normal middle ear function included normal immittance results including Type A tympanograms (Jerger, 1970) and the presence of 1 kHz ipsilateral acoustic reflexes at or below 100 dB HL (Silman and Gelfand, 1981).

Test materials were the FPT (Musiek, 1994) (DVA recording) (Noffsinger et al, 1994) and the DDT –2 (Musiek, 1983) (Auditec St. Louis, MO, recording) delivered at 50 dB HL. The FPT was delivered monaurally, whereas the DDT was delivered binaurally. Test and ear order were randomly selected. A total of 20 trials of DDT (2 pairs per trials) were used and 15 of the FPT per ear. For both tasks, a training component of three presentations was included. Task performance scores were expressed as percentage correct.

RESULTS

Raw percentage scores for both tasks (n = 180) are shown in Table 2. Results are displayed in single and 2 yr bands, to enable comparison with other published data.

Table 2. Initial Normative Results

		Left DDT		Right DDT		Left FPT		Right FPT	
Age (yr)	No.	Mean/SD	–2 SD	Mean/SD	–2 SD	Mean/SD	–2 SD	Mean/SD	–2 SD
7	13	81.8 ± 7.3	67.2	85.9 ± 7.9	70.2	78.2 ± 18.3	41.6	79.2 ± 18.7	60.5
8	52	85.2 ± 7.0	71.2	87.0 ± 8.1	70.8	69.8 ± 19.1	31.6	71.5 ± 18.7	34.1
7–8	65	84.7 ± 7.2	70.3	88.2 ± 8.1	72.0	71.5 ± 19.1	33.3	73.1 ± 18.8	35.5
9	12	85.3 ± 6.7	71.9	88.8 ± 5.6	77.6	77.8 ± 11.5	54.8	72.8 ± 12.6	47.6
10	49	91.2 ± 6.4	78.4	92.7 ± 6.8	79.1	87.0 ± 10.6	65.8	85.2 ± 9.9	65.4
9–10	61	90.1 ± 6.9	76.3	91.0 ± 6.7	77.6	85.1 ± 11.4	62.3	82.7 ± 11.6	59.5
11	4	89.4 ± 5.1	79.2	97.5 ± 2.9	91.7	82.0 ± 12.6	56.8	90.8 ± 8.3	74.2
12	50	93.3 ± 4.8	83.7	94.0 ± 4.6	84.8	91.1 ± 9.8	71.5	90.6 ± 9.1	72.4
11–12	54	93.0 ± 4.9	83.2	94.3 ± 4.6	89.7	90.4 ± 10.2	70	90.7 ± 8.9	72.9

Note: The results are also represented in 2 yr brackets and the –2 SD mark provided.

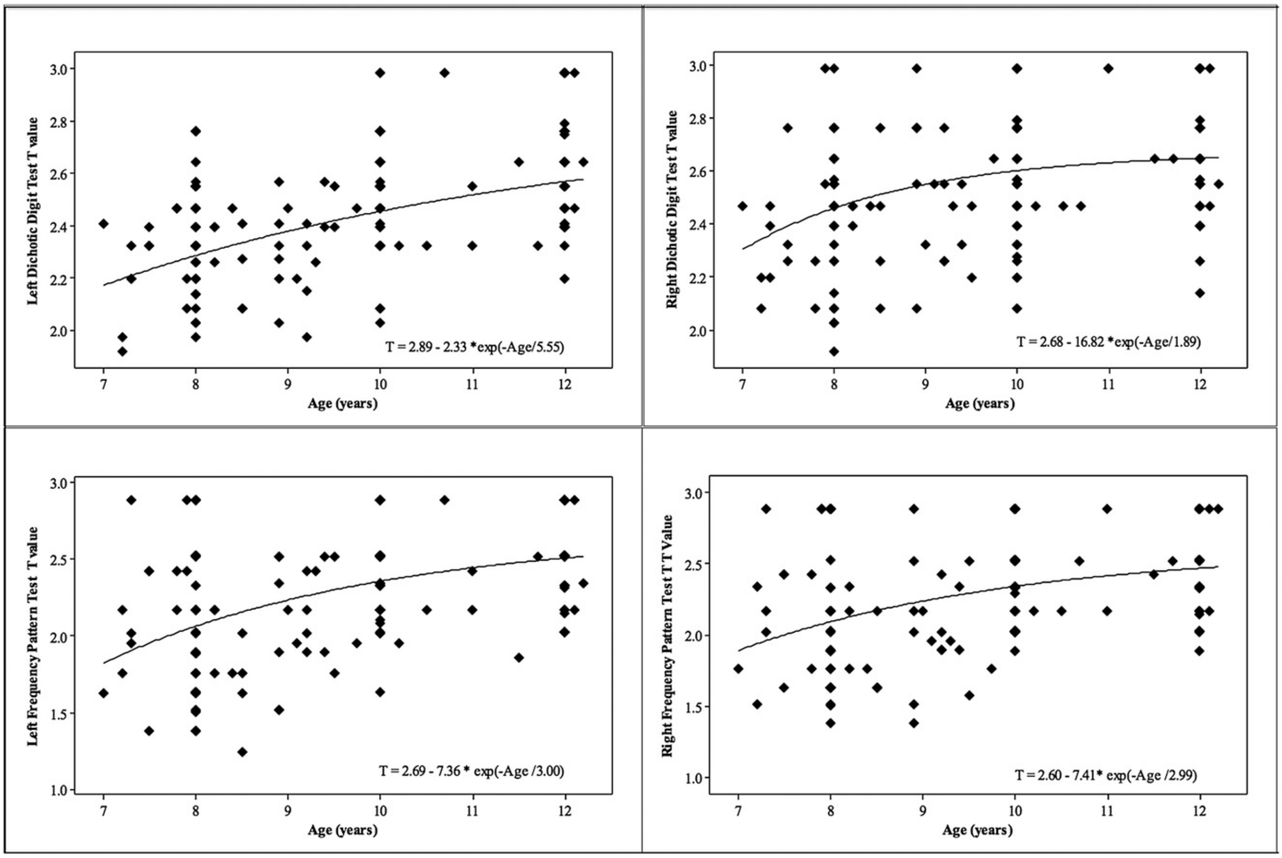

Figure 1. Non-linear regression results, showing T′ as a function of age.

Transformation of percentage results to give Z scores relative to age. The normative data are first arcsine transformed to normalize the variance, using Equation 1. These transformed scores are then regressed against age using an exponential equation, as shown below in Equation 3. The resulting regressions are shown in Figure 1. Note that age is expressed as a decimal value (e.g., 7.5 for 7 yr and 6 mo). Where the exact age of a child was unknown (missing data), the child's age was placed in the middle of the age bracket to which they were originally assigned:

$$T' = a + b^{*}\exp(-\text{age}/c), \tag{3}$$

where T′ is the estimated value of T, and a, b, and c are coefficients of the regression equation fitted to minimize the residual squared error about the regression line.

Finally, the squared values of the residual error around the exponential regression lines are regressed against age to estimate how the (squared) SD varies with age, as shown in Equation 4:

$$V' = f + g^{*}\text{age}, \tag{4}$$

where V′ is the estimated variance (i.e., squared SD), and f and g (Table 3) are the coefficients of the resulting regression equation.

The Z score for any individual percentage score can then be calculated using Equation 5, which is a specific form of Equation 2 using the mean score and SD expected for children of that age:

$$Z = [T - T']/\sqrt{V'}, \tag{5}$$

where T′ and V′ are calculated from Equations 3 and 4, respectively.

Any scores observed below 3 SD are excluded, and the process is repeated. Therefore, the results of one participant within this data set were excluded binaurally from the DDT.

The data are then examined to determine whether they form a normal distribution and are shown as the distributions in Figure 2. The raw data showed a skewed distribution verified by the Anderson Darling (AD) test of normality (left DDT: AD = 2.45, $p < 0.005$; right DDT: AD = 5.34, $p < 0.005$; left FPT: AD = 5.04, $p < 0.005$; right FPT: AD = 4.47, $p < 0.005$). After the transformation, the AD test of normality showed that the

Table 3. Regression Coefficients (f and g) for Use in Equation 4 for the Four Data Sets

Data Set	f	g
Left DDT	0.0381	0.000391
Right DDT	0.100	−0.00484
Left FPT	0.399	−0.0257
Right FPT	0.481	−0.034

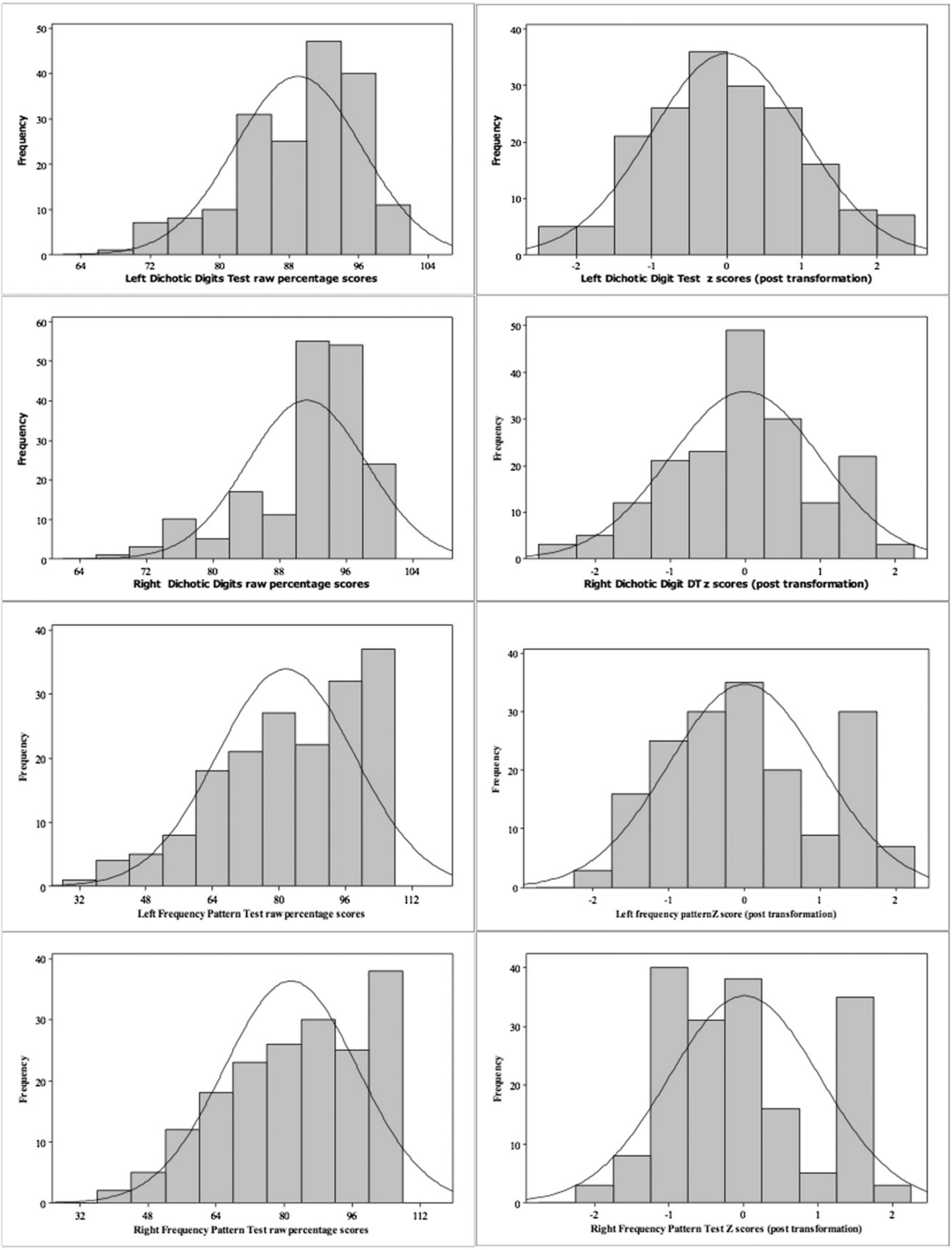

Figure 2. Distribution of DDT and FPT scores before (left) and after (right) transformation process.

dichotic scores did not differ significantly from normal (left DDT: AD = 0.49, $p = 0.22$; right DDT: AD = 0.79, $p = 0.04$). The FPT scores still differed significantly from normal but showed a lower degree of skew (left FPT: AD = 1.06, $p < 0.005$; right FPT: AD = 3.68, $p < 0.005$).

As an alternative to relying on the Z scores, the 2 SD cutoff scores resulting from the transformations described in this study can easily be calculated. First the lower cutoff scores are calculated in the transformed space by subtracting twice the age-appropriate SD (Eq. 4) from the age-appropriate means (Eq. 3). These transformed cutoff scores are then transformed back to percentages using the inverse of Equation 1, as given by Miller (1978), as shown in Equation 6:

$$p = 0.5\left\{1 - \mathrm{sgn}(\cos \mathrm{t})\left[1 - (\sin \mathrm{t} + (\sin \mathrm{t} - 1/\sin \mathrm{t})/\mathrm{n})^2\right]^{1/2}\right\}, \tag{6}$$

The resulting cutoff scores are shown as a function of age in Figure 3, along with the data on which they are based, and several previously published cutoff scores. Figure 3 also shows the corresponding limits that are 2 SDs above the mean. Because the score distributions still have some negative skew, even after arcsine transformation, the upper 2 SD limits were capped to the transform value corresponding to 100% before inverse transforming is performed.

DISCUSSION

The raw data before the transformation process are similar to those of Singer et al (1998) and Bellis (1996), with the mean and/or –2 SD point, on average, in agreement with the previous finding of Bellis and/or Singer et al. This finding suggests that the data under review are similar in form and distribution to other normative data sets. It would also suggest that differences among American, Australian, and New Zealand listeners have minimal impact on these specific tasks. Therefore, these normative results may be generalizable to populations other than the specific ones the data were obtained from.

The asymmetric distribution of the raw data is demonstrated clearly in Figure 2, with the distribution of scores being more negatively skewed than is true of a normal distribution. Consequently, when an SD is calculated based on the untransformed percentage scores, more

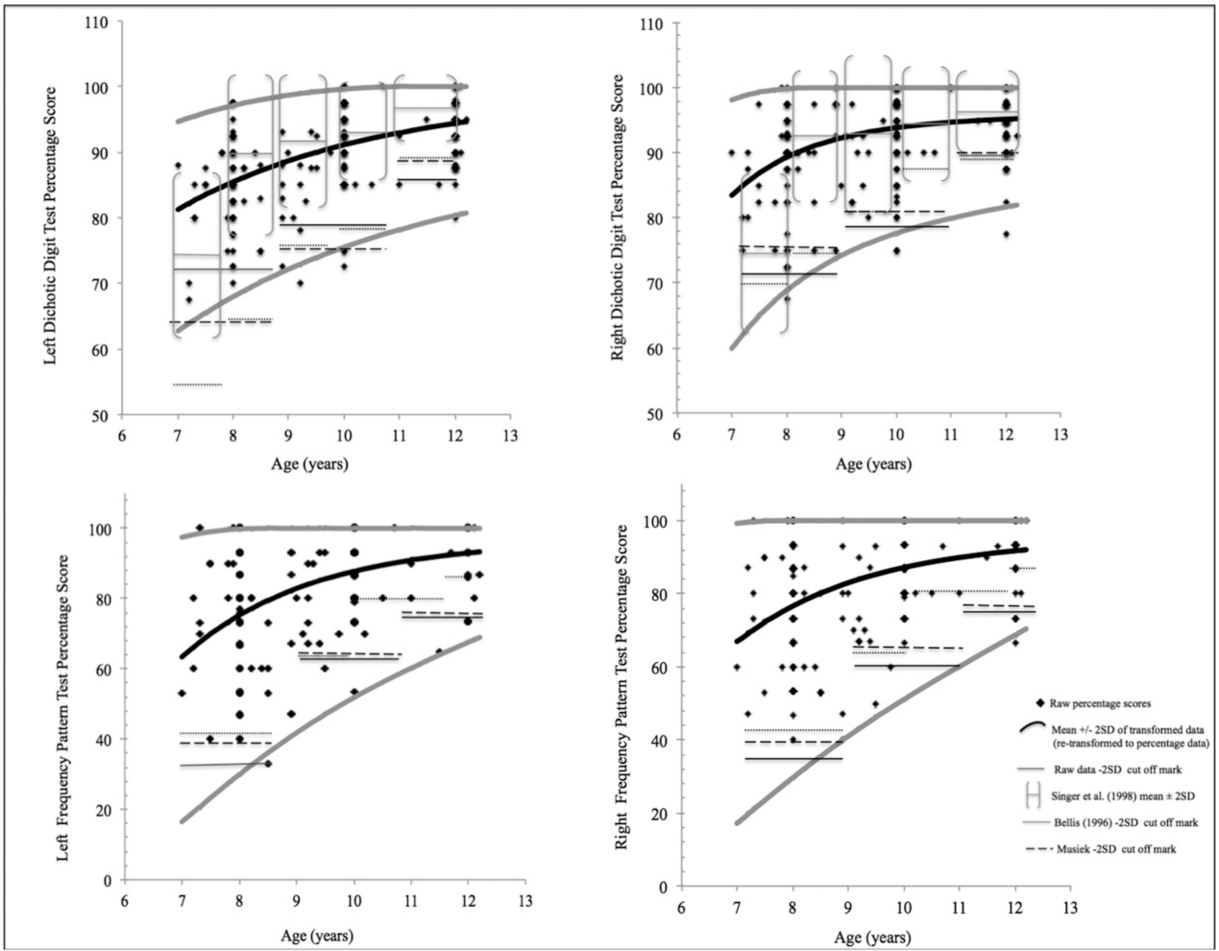

Figure 3. Comparison of percentage scores, means and standard deviations derived from transformed data, and published data.

than 2.5% of data points decrease more than 2 SDs below the mean. Furthermore, a value 2 SDs above the mean is typically more than 100%, which demonstrates the inappropriateness of using this statistic.

Also evident in Figure 2 is the normalization of the distribution of the same data after the transformation process. The data now show a –2 SD point that is considerably lower than would be obtained simply by calculating the means and SD of the raw percentage scores, in almost all ages, in all four data sets, as illustrated in Figure 3. The outline of the age-specific means and SDs of the transformed data represents the original data well, showing a similar spread of results. However, a poorer raw percentage score is now required to reach the –2 SD point. As a result, poorer scores have less negative Z scores than the untransformed results, whereas better scores show little change to resulting Z score before or after transformation.

The previously reported development of the skills with age is evident, confirming the need for age-specific normative data. Also seen are the changing rates of development, being greatest in the younger years (perhaps unsurprisingly), as demonstrated by the slope of the mean against age for both the raw data (Fig. 3) and the transformed data (Fig. 1). This steep rate of development confirms that a 2 yr, or even 1 yr, band of normative data is not appropriate. For example, Figure 3 shows that for the FPT, the –2 SD point varies by more than 15% (after retransformation to percentage data) across the 7–8 yr age range and again in the 8–9 yr age range. Therefore, we can improve the accuracy of normative data by taking this approach of smoothing, in an age-dependent manner, the results across age.

A recent emerging concept in the APD literature is the “degree of deficit” (Cameron and Dillon, 2008; Rosen et al, 2010; Boscariol et al, 2011; Dillon et al, 2012). Relying solely on a pass-or-fail criterion can oversimplify consideration of a child’s auditory processing ability. The use of a Z score, as generated by this interpretation of the data, allows for a clear, simple expression of the child’s performance. For example, the use of Z scores as a continuous variable will distinguish a child who just fails to meet some predetermined, arbitrary, cutoff score from a child who fails to meet the cutoff score by a large degree. Furthermore, it becomes clear that a child whose results are slightly poorer than the cutoff score has a very similar ability to a child whose results are only slightly better than the cutoff score. The underlying assumption is that the degree of disability experienced in real life will increase with the degree of deficit observed in the auditory processing test, with the more negative the Z score the greater the difficulty. This concept of degree of impairment is certainly one with which most health professionals are familiar. The clinical decision-making process is greatly strengthened by an awareness of the magnitude of the skill deficit present.

CONCLUSION

Formulae based on normative data are provided for tests of binaural integration and temporal patterning for clinical use. Raw percentage scores are transformed to age-specific Z scores (that have also accounted for the asymmetric distribution of the normative results). A simpler spreadsheet containing these formulae can be downloaded, or used as an online tool. Interested readers should visit the HEARnet site at http://www.hearnet.org.au/health-professionals-area/audiologists-speech-pathologists/resources/.

Acknowledgments. The authors thank Dr. Mridula Sharma, associate professor Gary Rance, associate professor Robert Cowan, and Mark Seeto for their comments on an earlier version of this manuscript.

REFERENCES

American Academy of Audiology 2010. Clinical Practice Guidelines: Diagnosis, Treatment and Management of Children and Adults with Central Auditory Processing Disorder. August 24, 2010. Available at: http://www.audiology.org/resources/documentlibrary/documents/capd%20guidelines%208-2010.pdf. Accessed August 3, 2014.

American Speech-Language-Hearing Association Task Force on Central Auditory Processing Concensus Development. (1996) Central auditory processing; current status of research and implications for clinical practice. *Am J Audiol* 5:41–52.

Baran J, Musiek F. (1999) Behavioural assessment of the central auditory nervous system. In: Musiek F, Rintelmann W, eds. *Contemporary Perspectives in Hearing Assessment*. Boston, MA: Allyn and Bacon.

Bellis TJ. (2003) *Assessment and Management of Central Auditory Processing Disorders in the Educational Setting: From Science to Practice*. 2nd ed. New York, NY: Delmar Learning, Thompson Learning.

Bellis TJ. (1996) *Assessment and Management of Central Auditory Processing Disorders in the Educational Setting. From Science to Practice*. San Diego, CA: Singular Publishing Group.

Bellis TJ, Ferre JM. (1999) Multidimensional approach to the differential diagnosis of central auditory processing disorders in children. *J Am Acad Audiol* 10(6):319–328.

Boscariol M, Guimarães CA, Hage SR, et al. (2011) Auditory processing disorder in patients with language-learning impairment and correlation with malformation of cortical development. *Brain Dev* 33(10):824–831.

Cameron S, Dillon H. (2008) The listening in spatialized noise-sentences test (LISN-S): comparison to the prototype LISN and results from children with either a suspected (central) auditory processing disorder or a confirmed language disorder. *J Am Acad Audiol* 19(5):377–391.

Dawes P, Bishop DV, Sirimanna T, Bamiou DE. (2008) Profile and aetiology of children diagnosed with auditory processing disorder (APD). *Int J Pediatr Otorhinolaryngol* 72(4):483–489.

Dillon H, Cameron S, Glyde H, Wilson W, Tomlin D. (2012) An opinion on the assessment of people who may have an auditory processing disorder. *J Am Acad Audiol* 23(2):97–105.

Emanuel DC. (2002) The auditory processing battery: survey of common practices. *J Am Acad Audiol* 13(2):93–117, quiz 118–119.

Iliadou V, Bamiou DE. (2012) Psychometric evaluation of children with auditory processing disorder (APD): comparison with normal-hearing and clinical non-APD groups. *J Speech Lang Hear Res* 55(3):791–799.

Jerger J. (1970) Clinical experience with impedance audiometry. *Arch Otolaryngol* 92(4):311–324.

Jerger J, Musiek F. (2002) On the diagnosis of auditory processing disorder. *Audiol Today* 14:19–22.

Kelly AS. (2007) Normative data for behavioural tests of auditory processing for New Zealand school children aged 7 to 12 years. *Aust N Z J Audiol* 29:60–64.

Miller JJ. (1978) The inverse of the Freeman-Tukey double arcsine transformation. *Am Stat* 32:138.

Musiek FE. (1983) Assessment of central auditory dysfunction: the dichotic digit test revisited. *Ear Hear* 4(2):79–83.

Musiek FE. (1994) Frequency (pitch) and duration pattern tests. *J Am Acad Audiol* 5(4):265–268.

Musiek FE, Gollegly KM, Ross MK. (1985) Profiles of types of central auditory processing disorders in children with learning disabilities. *Comm Disord Q* 9(1):43–63.

Noffsinger D, Wilson RH, Musiek FE. (1994) Department of Veterans Affairs compact disc recording for auditory perceptual assessment: background and introduction. *J Am Acad Audiol* 5(4):231–235.

Rosen S, Cohen M, Vanniasegaram I. (2010) Auditory and cognitive abilities of children suspected of auditory processing disorder (APD). *Int J Pediatr Otorhinolaryngol* 74(6):594–600.

Sharma M, Purdy SC, Kelly AS. (2009) Comorbidity of auditory processing, language, and reading disorders. *J Speech Lang Hear Res* 52(3):706–722.

Silman S, Gelfand SA. (1981) The relationship between magnitude of hearing loss and acoustic reflex threshold levels. *J Speech Hear Disord* 46(3):312–316.

Singer J, Hurley RM, Preece JP. (1998) Effectiveness of central auditory processing tests with children. *Am J Audiol* 7: 73–84.

Studebaker GA. (1985) A "rationalized" arcsine transform. *J Speech Hear Res* 28(3):455–462.

Wilson WJ, Heine C, Harvey LA. (2004) Central auditory processing and central auditory processing disorder: Fundamental questions and considerations. *Aust N Z J Audiol* 26:80–93.