25

UNDERSTANDING MELODIC DICTATION VIA EXPERIMENTAL METHODS

David John Baker

Establishing the Problem

Melodic dictation is a cognitively demanding process that requires individuals to be able to hear a melody and then transcribe what they hear without any external reference to aid them. It is a staple of many aural skills curricula because the mastery of melodic dictation brings with it a deeper understanding of goals central to the aural skills classroom. Despite its centrality and near ubiquity in music education, accounting for all the variables that will inevitably cause a student to perform well on a melodic dictation is cumbersome.

Not only is melodic dictation a complex, cognitive process requiring elaborate mental choreography (G. Karpinski, 1990; G. S. Karpinski, 2000), but there is also variability in both individual ability (Cowan, 2014; Jakubowski, Müllensiefen, et al., 2017) and the structure of the melodies themselves (Harrison et al., 2017; Müllensiefen & Halpern, 06/2014) that researchers have demonstrated affects how people remember musical materials. With so many factors to account for, it is no wonder that some pedagogues have expressed an interest in beginning to standardize certain benchmarks for students as they progress through their aural skills development (Paney & Buonviri, 01/2014). From a pedagogical standpoint, questions of standardization help pedagogues understand what can reasonably be expected of students in order for them to be fairly assessed. Having a clearer understanding of what factors might contribute to an individual's ability to perform well in melodic dictation is also important so that assessments do not inadvertently measure factors beyond a student's control.

Often aural skills pedagogues rely on their expert opinion when it comes to assessing ability in aural skills, but research from decision-making sciences demonstrates how expert opinion often leads to inferior objective assessments of individual cases due to unconscious biases (Kahneman, 2012; Meehl, 1954). As a pedagogical community, it is important to understand what factors contribute to an individual's ability to take melodic dictation, not only to ensure fairness in assessing a student's work, but additionally because understanding what factors contribute to an individual's ability to take melodic dictation allows pedagogues to better explain how students can improve and grow as they progress through their aural skills education.

With so many variables at both the individual and musical level, how does one even begin to keep track of all these moving parts? One way to help organize and systematize research in aural skills research is to borrow tools from the sciences, since questions about aural skills ability are fundamentally questions about music perception. While some might see using such blunt reductions of melodies and considerations of factors typically outside the scope of the classroom as incompatible to our pedagogical goals as teachers, the reason for doing this is to see what insights can be gleaned to help teaching rather than centering these research methods as didactic tools. Specifically, I borrow from computational musicology to help discuss the sonic elements related to melodic dictation, and borrow from cognitive psychology in order to discuss individual differences. This chapter continues to build on the past three decades of research bridging the gap between music cognition and aural skills. Creating an accessible resource for aural skills pedagogy serves as a way to continually bridge the gap between aural skills pedagogy and the world of music cognition (Brown, 2000–2001; Butler & Lochstampfor, 1993; David Butler, 1997; G. S. Karpinski, 1990; G. S. Karpinski, 2000).

Melodic Dictation: A Lot to Think About

Before diving into how methods from the sciences might help in understanding the complex process that is melodic dictation, let me illustrate why such systemization is needed. Imagine the following scenario: a second-year undergraduate oboe player without absolute pitch, majoring in music, must dictate a syncopated, arpeggiating melody in C♯ minor played in the lower register of the piano at a quick tempo after one playing of the melody.

From the teacher's standpoint, it might be easy to imagine what you, as an instructor, could do in order to make this task easier for the student. For example, quick and easy fixes that might make this more doable would be to increase the number of hearings the student is allowed, slow down the tempo, change the octave that the melody is played in, and tell the student they are allowed to write down the melody in a key other than C♯ minor. Taking full advantage of the fact that this is a thought experiment, other variables in this situation that might make the task even easier, but are more imaginative, would be to give the student absolute pitch and maybe turn them into a graduate student who has taught a few semesters of aural skills. Of course the first set of changes are more practical than being able to magically give someone absolute pitch or add on years of musical experience, but these changes exemplify that external factors are at play. Any changes fundamentally modify the exercise as a whole, but this fictitious thought experiment serves to demonstrate that with melodic dictation, there are many factors that presumably affect performance, with some affecting more than others. Acknowledging that there are many factors that would contribute to this, the following two questions are those an aural skills researcher looking to understand this process might have:

1. How does one keep track of the many factors that affect performance?
2. How is it then possible to take the qualitative features described earlier and turn them into something not dependent on the relative experience of the individual making this judgment?

In the next section, I first address the question of how to keep track of so many moving parts, and then I answer the second question by introducing how tools from cognitive psychology and computational musicology can help remove this relative ambiguity.

Keeping Track of It All

In order to better organize specific factors thought to contribute to melodic dictation, I present Figure 25.1, which provides a taxonomy to help organize the many factors that contribute to an individual's ability to take melodic dictation originally presented in David John Baker's doctoral dissertation (Baker, 2019).

The taxonomy in Figure 25.1 organizes the factors thought to contribute to an individual's performance on melodic dictation, or any form of musical memory task, by initially splitting the factors into two parts: individual and musical features. I put forward this taxonomy to help answer my first question of how to best keep track of these many factors.

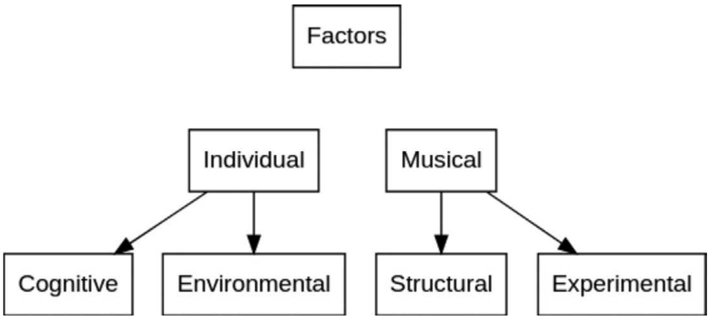

Figure 25.1 Taxonomy of factors contributing to individual performance.
Source: Baker (2019).

As discussed in Baker (2019), individual factors from the taxonomy in Figure 25.1 are those that are different between people and do not take into account any sort of sonic features in the context of melodic dictation. The individual factor bifurcates into both cognitive and environmental features. Cognitive factors refer to any aspect of an individual that deals with the process of thinking. Examples of these would include if the individual has absolute pitch, his or her working memory capacity, or could even be age if used as a proxy for development. These factors are typically beyond the conscious control of the individual. Environmental factors, on the other hand, describe aspects of an individual that reflect prior experiences such as when he or she began taking music lessons, what instrument he or she primarily practices, or could extend to how much money has been invested into musical training. The categories are not meant to be mutually exclusive but rather a descriptive framework to help organize thinking before employing the scientific tools we will see later.

Musical factors encapsulate both structural and experimental aspects of a melodic dictation. Structural aspects of the melody are features of the melody that remain invariant when written down in musical notation that only capture changes in pitch over time. Experimental factors refer to how the melody is performed: the timbre of the melody, how fast the melody is played, and how many times it is played. Again, the categories in this taxonomy are not meant to be mutually exclusive, and this taxonomy favors European, Western conceptualizations of music reflecting that melodic dictation is a European, Western conception.

Readers familiar with other literature on melodic dictation will notice the taxonomy in Figure 25.1 accounts for variables beyond those put forward by Karpinski (2000), whose didactic model focuses solely on the mental choreography required of the individual and does not include musical features thought to contribute to the model. Figure 25.1 intends to provide not only a means to conceptualize the factors of interest but, importantly, a way for empirical researchers to organize what variables might contribute to an individual's ability to take melodic dictation.

Taxonomizing has many benefits. The first is that it allows researchers to answer the first question asked earlier and consequently organize research on the many factors into more tangible mental categories. Having language to keep track of all this allows the researcher to focus on what he or she is interested in and accounts for variables that might confound what is hoped to be better understood. The second benefit is that, given a mental model of separating factors into separate variables, this consequently allows aural skills researchers to begin addressing the second question, so that musical features can be operationalized as variables to be manipulated in an experimental setting. Being able to operationalize – meaning deciding how something abstract will be measured – allows for several important benefits which ultimately lay the groundwork and provide the rationale as to why tools from computational musicology and cognitive psychology can help research in aural skills. Next, we explore why this operationalization is key to being able to incorporate both work from computational musicology and cognitive psychology.

As many people who teach music are aware, it is very difficult to talk about melodies without relying on the often jargon-heavy language provided by music theory and analysis. There is not, nor should there be, any sort of deterministic language to describe musical structures, so often those in charge of teaching will let their own subjectivities guide what is needed for classroom instruction. Due to this general relativity in discussing musical structures, what one teacher might deem as 'easy-to-dictate' for a certain group of students they are responsible for teaching might also be described as 'difficult-to-dictate' by another teacher with responsibilities to a different population of students. In daily pedagogical practice, this relativity is not often a problem since teachers use their own judgment to suit the needs of their classroom. The problem with this is that relative judgments of difficulty will not generalize beyond the immediate classroom, making it more difficult to share information with others.

Instead of attempting to reconcile subjective differences and reach an agreement between many individuals, we can instead borrow ideas from the field of computational musicology to get an objective proxy for difficulty that will be invariant regardless of who makes the judgment. For example, it is much easier to come to a consensus about how many notes are in a melody and at what tempo a melody is played (in beats per minute) than to mark a melody with a relative judgment such as 'suitable for a first-year student' or 'of medium difficulty.' In addition to tethering features of a melody to something quantitative and objective, using these types of measures allows the field of aural skills to be able to engage with findings from cognitive psychology and computational musicology to help inform the direction of a research program.

In the next section, we conduct a brief exploration through literature in computational musicology and cognitive psychology to show how empirical tools can help bring a clearer understanding of factors that contribute to melodic dictation. Being able to quantify factors we presume contribute to how an individual performs in melodic dictation – our dependent variable of interest – opens up the possibility of statistically modeling these relationships, bringing the community one step closer to understanding the inner workings of this process.

What the Scientific Literature Has to Offer

Using the taxonomy from Figure 25.1 as our guide, we can now investigate literature that can inform future work on melodic dictation. I begin with exploring individual features and then explore work looking at musical features.

Considering both cognitive and environmental individual factors simultaneously, it is not difficult to find studies that report that both musical training and preexisting individual cognitive differences as factors that can be used as successful predictors in tasks of musical perception. For example, in a meta-analysis by Talamini and colleagues, the authors found that musicians tended to outperform their non-musical counterparts on many memory tasks, especially if that task had tonal elements (Talamini et al., 2017). This finding corroborates earlier ideas put forward by Berz, who noted that many tasks that are used in the world of music perception share many similarities with the concept of working memory (Berz, 04/1995). This finding is important in regard to any empirical work in melodic dictation because recently some scholars like Chenette (2019) have argued in favor of reframing work in aural skills around concepts regarding working memory.

Chennette centers these ideas in his own work and is corroborated by researchers in music perception who have shown links between the ability to remember melodic information and performance on musical memory tasks (Colley et al., 08/2017; Nichols et al., 06/2018). It follows that if concepts like working memory capacity are instrumental in understanding tasks of musical memory, as suggested by both Berz and Chennette, then presumably variation in working memory capacity would be something to consider in any empirical investigation of performance in something like melodic dictation. Additionally, there has been a considerable amount of theoretical work devoted to studying working memory on which researchers interested in applying those findings to in a musical context can base their work (Cowan, 2005).

Likewise, while a large amount of literature has shown mixed results linking cognitive ability and general musical training (Swaminathan & Glenn Schellenberg, 2018), recent meta-analyses have noted serious problems in this program of research, even calling for its dismissal after no effects have been demonstrated when controlling for quality of study (Sala & Gobet, 2020). In surveying this literature, aural skills researchers are able to save time and resources and investigate hypotheses that are more likely to be correct.

Turning now to factors on the musical side of the taxonomy, researchers in music psychology have demonstrated that computationally derived summary features of melodies have been successful in predicting performance on melodic tasks. While many of these summary features do not yet have a track record of mapping to features in melodic dictation tasks, computationally derived features have been used to predict court case decisions in cases of musical plagiarism (Müllensiefen & Pendzich, 2009), historical data predicting the position of songs from the Beatles' album *Revolver* (Kopiez & Müllensiefen, 2011), memory for melodies in musical recall experiments (Harrison et al., 2017; Müllensiefen & Halpern, 06/2014), memory for earworms (Jakubowski, Finkel, et al., 2017; Williamson & Müllensiefen, 2012), and memorability of pop music hooks (Van Balen et al., 2015). There has been research suggesting that the structural features of melodies, when distilled into a composite measure of complexity (David J. Baker & Müllensiefen, 2017; Harrison et al., 2017), can also be predictive of performance on tasks involving musical memory.

In each case listed here, the degree that a certain melody was remembered was able to be predicted above chance accuracy using an assemblage of features from the FANTASTIC toolbox (Müllensiefen, 2009). If successful in prediction settings in the music perception literature, there is no reason these same features could not be further extended to melodic dictation paradigms as in work from Baker, Monzingo, and Shanahan (Baker, Monzingo, & Shanahan, 2018).

Experimental Walk-Through

Having now explored just a subset of the many variables that research in computational musicology and cognitive psychology have shown to be linked to melodic memory, how would one begin a line of empirical research in order to investigate this further? What pitfalls should a practitioner who is well versed in aural skills teaching, but not empirical methods, be aware of? In this next portion of this chapter, I take a high-level walk-through of ideas and concepts someone looking to start a research project might consider before embarking on his or her own program of research in this area.

The first temptation a researcher might be drawn to, having read literature on factors affecting musical memory, is to create a giant list of every variable of interest to use as a predictor, in hopes of using them to understand performance on melodic dictation. This approach is discouraged, as the inclusion of too many variables in a prediction leads to overfitting. While there is no way to explain the intricacies of experimental design and data analysis within the scope of this chapter, there are multiple resources interested researchers can use to supplement their education (McElreath, 2020). As collecting data for music psychology experiments has historically run the risk of claiming to discover true findings when none exist for many reasons (Frieler et al., 2013), research on melodic dictation is no different in this regard. For these reasons, the first important task for the researcher is to pick variables that are ideally theoretically viable and pedagogically of interest.

For example, we might note from the aforementioned that work by Eerola and colleagues (Eerola, 2016) reported that note density has shown to be a predictor in experiments of musical memory. From a pedagogical standpoint, this is a variable of interest since it is related to questions about working memory capacity and chunking (Chenette, 2019) and is calculated by dividing the number of notes by the time period it takes to sound them. There might be other variables that presumably affect this on the musical level, so perhaps, getting more abstract, we could select a measure of interval entropy since there has also been work showing how measures of information content, a mathematical formalization of expectedness, can be predictive of memory on musical memory tasks (Pearce, 07/2018). Researchers should be cautious of the fact that both of the hypothetical variables selected are not completely independent of one another. This colinear relationship illustrates something that has been noted in the literature going back to Taylor and Pembrook (1983) within musical features investigating musical recall.

It is possible to include both variables in the model, but as more are added, the addition of multiple colinear variables in our eventual statistical model will lead to instability (Harrell Jr., 2015). Again, some researchers in the field of music perception have addressed this statistical problem using certain forms of data reductive techniques (David J. Baker & Müllensiefen, 2017; Harrison et al., 2017), but this comes at the price of having to interpret the final output at a layer of abstraction that does not directly map to the initial variables of interest. While an almost infinite number of metrics could be used to operationalize features on the musical side of the taxonomy, software for computing these features exists such as Müllensiefen's FANTASTIC tool box that in this case is capable of computing both note density and interval entropy (Müllensiefen, 2009).

In addition to measuring at the musical level, it is also possible to keep track of variables at the individual level. From the fictitious example in the introduction, it would be reasonable from an intuitive standpoint to be aware of variables like musical training or number of weeks of aural skills education, but, as noted earlier in the research from music psychology, there is literature to suggest the importance of applying these individual differences to our final statistical models. Presumably a considerable amount of variation in our responses can be explained by just the variation within the baseline ability of an individual. For example, we may include the two aforementioned variables (number of weeks of aural skills classes taken or the age that an individual started music lessons).

In both the case of the musical and individual features, from an experimental standpoint, each of these variables would have to be collected and tracked for every individual taking part in a study on melodic dictation. In addition to these variables of interest, past experience would also remind aural skills instructors that variables such as the tempo at which each melody is to be played, the timings between hearings, and how many hearings are included will inevitably affect performance on melodic dictation.

Given the number of variables at play, ideally this can be done with an automated process. One final caveat before moving on to the data collection is that when selecting melodies, researchers need to ensure that whatever melodies are chosen, it is important to select melodies that are neither too easy nor too difficult so that participants will neither score perfectly nor be unable to complete the task. While a perfect score would be something to celebrate in a pedagogical context, having many individuals able to completely ace a melodic dictation creates something referred to as a ceiling effect, making it impossible to learn anything since the information from a scientific point of view comes from the variation in responses. Participants should be informed that the melodic dictation they will partake in for this experimental context is very different from what they experienced before; they should not be able to do it perfectly. This information might allay stress for some participants, but in my experience, some participants who pride themselves in their aural skills ability take this as a challenge.

The experiment should be administered under conditions that are controlled and documented. While some aural skills researchers may be able to exactly reproduce performances on the piano (or their instrument of choice) with strict adherence to tempo using a metronome and clock to time between hearings, I suggest that people incorporate software in order to have more control and to remove stress from the researcher conducting the experiment. While the setup of each data collection will change depending on its unique needs, all of the variables of importance indicated earlier should be tracked. An experimental melodic dictation can often look very much like it is done within the classroom but carried out with a computer. It is also of paramount importance to consult with your institution's board of ethics before collecting any sort of data that will be used for research purposes.

Upon following a method of data collection, the next thing that is needed is to be able to score each melodic dictation. Unlike classroom marking, the scoring of each melody should be carefully indicated beforehand and documented. Ideally this process will involve at least independent grading by a second person so that some metric of inter-rater reliability can be calculated to quantify the level of discrepancy between raters. If raters are able to achieve a high degree of consistency between themselves, composite scores can be used as the dependent variable in statistical modeling.

The last step in this process would then be to statistically model your results. While there are numerous pedagogical textbooks on statistical modeling that you might explore, and even designs specifically about music (Baumgartner, 2019), in this final walk-through of the process I will instead focus on what can be gleaned from a statistical analysis rather than how to conduct one.

In the more complex designs like the one we are working with here, where we track both musical and individual features, there are many strategies and resources that exist for those looking to learn more (Harrell Jr., 2015; McElreath, 2020). Typically, what is of interest in this context of statistical modeling is to know how good your statistical model is: given what you thought you would be good at predicting, how well are you able to actually predict what you set out to do in the first place? In this example, we plan to use measures of note density, interval entropy, and a measure of musical training as our independent variables to predict our dependent variable, which is the final score.

There are multiple ways in which one could model this. Before carrying out any sort of data collection, it is always best to consult with a statistician. Further recommendations at this stage would be to pre-register any hypotheses you have (Nosek et al., 2018) and to plan to make both your data and analysis accessible to others so they are able to reproduce your results. Of interest in this fictitious analysis are two elements. The first is to find out how good your model is. The second is the ability to investigate the degree to which each variable contributes to your entire model's predictive ability. Both of these elements are typically found in a regression table like the one presented in Table 25.1.

Table 25.1 Mixed effects modeling of melodic dictation data.

		Percent Rev
Predictors	*Estimates*	*CI*
(Intercept)	100.33	100.16−100.50
auralWeeksTaken	−0.00	−0.00−0.00
note.dens	−0.13	−0.14−−0.11
i.entropy	−0.55	−0.78−−0.32
Random Effects		
σ^2	0.01	
τ_{00} subjectNo	0.11	
τ_{11} subjectNo.i.entropy	0.43	
ρ_{01} subjectNo	−0.89	
ICC	0.72	
$N_{\text{subjectNo}}$	41	
Observations	704	
Marginal R^2 / Conditional R^2	0.439 / 0.845	

Source: Data collected in Baker (2019).

Again, while there is no way to explain all aspects of a statistical analysis in a single chapter, I highlight here what is of importance. The first element requires us to inspect how well the model fits the data.

While there are many metrics to compare models, this example focuses on using R^2: a measure that is defined by the proportion of the variance in the dependent variable – our final score – that is predictable from the independent variables: note density and interval entropy. In the fictitious data reported in Table 25.1, the marginal R^2 value is 0.439, which is how much of the data we can explain with the variables we entered into the model. The conditional R^2 is 0.845, which also includes variance the model is able to capture that is due to just how participants differ. The R^2 value typically goes between 0.00 and 1.00. A value of 0.00 would indicate that we are basically unable to learn anything from our model other than it was a bad model, with a 1.00 suggesting we have overfit our data.

Seeing as this model is able to predict about 44% of the data, it means that 56% is left unexplained, but we know a fair bit of this has to do just with differences in the baseline abilities of our 41 participants. Of the results we can explain with this model, we can then inspect the next question of the degree to which each of our variables affects our results. This information is found in the coefficient column of Table 25.1. We can interpret these values as meaning that for every 1 unit increase in the interval entropy, we can expect the final score to decrease by -0.55, all while accounting for the effects of interval entropy and weeks of aural skills training. The same logic can be applied to the other variables in the model. This finding should align with intuitions. We would imagine that as note density goes up, the final average score should go down as reflected by the negative coefficient. While most aural skills instructors may have been able to intuit the direction of this relationship, what this statistical analysis shows is both the magnitude and amount of error associated with this estimation. As this process reflects a stochastic, as opposed to deterministic process, this number will invariably change a little bit from one study to the next, but the information gleaned from this small analysis can be used to inform future work building the bridge between music cognition and aural skills pedagogy.

Discussion

After even a high-level walk-through of everything required to obtain this empirical data, some researchers might be put off by the amount of information gained, given how much work is required to obtain these estimations. Writing all of this out in a narrative shows how much work is required and possibly explains why such scant literature exists and why the bridge between music perception and aural skills is often in need of being built. Although the process is slow, what would many researchers adopting these methods enable?

The first, and probably most important result from developing a strong empirical attitude toward melodic dictation research would be a better understanding of the bounds of what can be expected from students at various points in their education. If reliable standards were obtained, students and teachers would be able to have more realistic yardsticks of achievement for assessment on the global level. Individually, this would also allow for more of a fine-grained approach at establishing difficulty, making it easier for students to move linearly, as opposed to the large jumps we often ask of students accidentally.

A second side effect of adopting these methods more generally would be more sharing of knowledge across the aural skills community and more collaboration. Not only would this be possible within aural skills, but because many skill sets are required between experimental design, computational analyses of melodies, and statistical analysis, work in this area could bridge more gaps.

Third, this dialogue using language of aurals skills will lead to unexpected findings and new lines of questions within aural skills. This volume shows growing interest in aural skills, and now is the time to reimagine what aural skills research could be.

This chapter has covered a lot of ground in showing this process from start to end. There are three general benefits that the reader will hopefully take away. The first is the taxonomy in Figure 25.1 which helps organize the many factors that are at play in this complex process. The second benefit introduces how tools from the sciences might provide for a more objective way to talk about difficulty in melodies. I introduced the ideas of features from the computational musicology literature – numerical features that could be a numerical proxy stand-in for melody – and how they might be helpful. Lastly, the walk-through of points to consider when running an experiment and partial data analysis demonstrates how those who are interested in the music cognition side of the bridge, but live on aural skills side, can begin to walk back and forth.

References

Baker, D. J. (2019). *Modeling Melodic Dictation* [Louisiana State University]. https://digitalcommons.lsu.edu/gradschool_dissertations/4960/

Baker, D. J., Monzingo, E., & Shanahan, D. (2018). Modeling Aural Skills Dictation. In Richard Parncutt, Sabrina Sattmann, Christine Beckett, Eldad Tsabury, Isabel Martinez, & Emery Schubert (Eds.), *Proceedings of the 15th International Conference on Music Perception and Cognition – Graz*, Austria; Montreal, Canada; La Plata, Argentina; Sydney, Australia; 23–28 July 2018. Retrieved July 30, 2020, from www.researchgate.net/profile/David_Baker45/publication/326816344_Modeling_Aural_Skills_Dictation/links/5b64a3c9aca2724c1f201e9f/Modeling-Aural-Skills-Dictation.pdf

Baker, D. J., & Müllensiefen, D. (2017). Perception of Leitmotives in Richard Wagner's Der Ring des Nibelungen. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00662

Baumgartner, C. M. (2019). Statistics in Music Education Research: A Reference for Researchers, Teachers, and Students. *Music Education Research*, *21*(2), 210–211.

Berz, W. L. (04/1995). Working Memory in Music: A Theoretical Model. *Music Perception: An Interdisciplinary Journal*, *12*(3), 353–364.

Brown, M. (2000–2001). Music Theory, Music Cognition, and the Case of the Iguanodon's Thumb. *Intégral*, *14/15*, 72–86.

Butler, D. (1997). Why the Gulf between Music Perception Research and Aural Training? *Bulletin of the Council for Research in Music Education*, *132*. Retrieved from www.jstor.org/stable/40375331

Butler, D., & Lochstampfor, M. (1993). Bridges Unbuilt: Comparing the Literatures of Music Cognition and Aural Training. *Indiana Theory Review*, *14*(2), 1–17.

Chenette, T. K. (2019). Reframing Aural Skills Instruction Based on Research in Working Memory. *Journal for Music Theory Pedagogy*, *18*.

Colley, I. D., Keller, P. E., & Halpern, A. R. (08/2017). Working Memory and Auditory Imagery Predict Sensorimotor Synchronisation with Expressively Timed Music. *The Quarterly Journal of Experimental Psychology*, *71*(8), 1781–1796.

Cowan, N. (2005). *Working Memory Capacity*. East Sussex, England: Psychology Press.

Cowan, N. (2014). Working Memory Underpins Cognitive Development, Learning, and Education. *Educational Psychology Review*, *26*(2), 197–223.

Eerola, T. (2016). Expectancy-Violation and Information-Theoretic Models of Melodic Complexity. *Empirical Musicology Review: EMR*, *11*(1), 2.

Frieler, K., Müllensiefen, D., Fischinger, T., Schlemmer, K., Jakubowski, K., & Lothwesen, K. (2013). Replication in Music Psychology. *Musicae Scientiae: The Journal of the European Society for the Cognitive Sciences of Music*, *17*(3), 265–276.

Harrell, F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. New York City, NY: Springer.

Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying Modern Psychometric Techniques to Melodic Discrimination Testing: Item Response Theory, Computerised Adaptive Testing, and Automatic Item Generation. *Scientific Reports*, 7(1), 3618.

Jakubowski, K., Finkel, S., Stewart, L., & Müllensiefen, D. (2017). Dissecting an Earworm: Melodic Features and Song Popularity Predict Involuntary Musical Imagery. *Journal of Aesthetics, Creativity, and the Arts*, *11*(2), 112–135.

Jakubowski, K., Müllensiefen, D., & Stewart, L. (2017). A Developmental Study of Latent Absolute Pitch Memory. *Quarterly Journal of Experimental Psychology*, *70*(3), 434–443. https://doi.org/10.1080/17470218.2015.1131726

Kahneman, D. (2012). *Thinking, Fast and Slow*. New York City, NY: Penguin Books.

Karpinski, G. S. (1990). A Model for Music Perception and Its Implications in Melodic Dictation. *Journal of Music Theory Pedagogy*, *4*(1), 191–229.

Karpinski, G. S. (2000). *Aural Skills Acquisition: The Development of Listening, Reading, and Performing Skills in College-level Musicians*. Oxford: Oxford University Press.

Kopiez, R., & Müllensiefen, D. (2011). Auf der Suche nach den 'Popularitätsfaktoren' in den Song-Melodien des Beatles-Albums Revolver: Eine computergestützte Feature-Analyse [In Search of Features Explaining the Popularity of the Tunes from the Beatles Album Revolver: A Computer-assisted Feature Analysis]. *Musik Und Popularitat*, 207–225.

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. Boca Raton, FL: CRC Press.

Meehl, P. E. (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis, MN: University of Minnesota Press.

Müllensiefen, D. (2009). *Fantastic: Feature Analysis Technology Accessing Statistics* (In a Corpus): Technical Report v1.5.

Müllensiefen, D., & Halpern, A. R. (06/2014). The Role of Features and Context in Recognition of Novel Melodies. *Music Perception: An Interdisciplinary Journal*, *31*(5), 418–435.

Müllensiefen, D., & Pendzich, M. (2009). Court Decisions on Music Plagiarism and the Predictive Value of Similarity Algorithms. *Musicae Scientiae: The Journal of the European Society for the Cognitive Sciences of Music*, *48*, 257–295.

Nichols, B. E., Wöllner, C., & Halpern, A. R. (06/2018). Score One for Jazz: Working Memory in Jazz and Classical Musicians. *Psychomusicology: Music, Mind, and Brain*, *28*(2), 101–107.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The Preregistration Revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606.

Paney, A. S., & Buonviri, N. O. (01/2014). Teaching Melodic Dictation in Advanced Placement Music Theory. *Journal of Research in Music Education*, *61*(4), 396–414.

Pearce, M. T. (07/2018). Statistical Learning and Probabilistic Prediction in Music Cognition: Mechanisms of Stylistic Enculturation: Enculturation: Statistical Learning and Prediction. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395.

Sala, G., & Gobet, F. (2020). Cognitive and Academic Benefits of Music Training with Children: A Multilevel Meta-analysis. *Memory and Cognition*. https://doi.org/10.3758/s13421-020-01060-2

Swaminathan, S., & Glenn Schellenberg, E. (2018). Music Training and Cognitive Abilities: Associations, Causes, and Consequences. In Michael H. Thaut and Donald A. Hodges (Eds.), *The Oxford Handbook of Music and the Brain*. Oxford: Oxford University Press.

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians Have Better Memory than Nonmusicians: A Meta-analysis. *PloS One*, *12*(10), e0186773.

Taylor, J. A., & Pembrook, R. G. (1983). Strategies in Memory for Short Melodies: An Extension of Otto Ortmann's 1933 Study. *Psychomusicology: A Journal of Research in Music Cognition*, *3*(1), 16–35.

Van Balen, J., Burgoyne, J. A., & Bountouridis, D. (2015). Corpus Analysis Tools for Computational Hook Discovery. *Proceedings of International Society for Music Information Retrieval*, *15*.

Williamson, V. J., & Müllensiefen, D. (2012). Earworms from Three Angles: Situational Antecedents, Personality Predisposition and the Quest for a Musical Formula. *Proceedings from the 12th International Conference on Music Perception and Cognition*, 1124–1132.