UvA-DARE (Digital Academic Repository)

The Role of Absolute Pitch Memory in the Oral Transmission of Folk Songs

Olthof, M.; Janssen, B.; Honing, H.

[Link to publication](#)

The Role Of Absolute Pitch Memory
In The Oral Transmission of Folksongs

MERWIN OLTHOF
MCG, ILLC, University of Amsterdam

BERIT JANSSEN[1]
Meertens Instituut

HENKJAN HONING
MCG, ILLC, University of Amsterdam

ABSTRACT: Absolute Pitch (AP) is the ability to identify or produce isolated tones in the absence of contextual cues or reference pitches. While AP is thought to differ from other human abilities in its bimodal distribution (Takeuchi & Hulse, 1993) – either you have it or you do not – recent evidence suggests that memory for absolute pitch in a melody is actually widespread (Schellenberg & Trehub, 2003). In the current project the Dutch collection of historic audio recordings, *Onder de Groene Linde (Grijp, 2008)*, is used as a source to explore the potential role of AP in the memory of songs transmitted in oral traditions. Since the melodies in this database are grouped by *tune family* and are available as sound files, they can serve as empirical support for the Absolute Pitch Memory (APM) hypothesis, predicting that these tunes are memorized and transmitted over time and geographical location based on their absolute pitch height. To this end, between- and within-tune family analyses were performed. In the between-tune family analysis, two tune families showed significant inter-recording tonic pitch consistency. The within-tune family analysis further substantialized that effect, while controlling for possible factors of variance such as gender, geographical origin, and lyrics. Together, the results are taken as empirical support that APM plays a significant role in the oral transmission of folksongs.

INTRODUCTION

WHILE the ability to instantly identify and label an isolated tone as being a particular note in the tonal system is very rare (Takeuchi & Hulse, 1993), research suggests that memory for absolute pitch information is, in fact, widespread (e.g., Levitin, 1994; Schellenberg & Trehub, 2003). Expanding on these earlier studies, this study analyzed pitches of Dutch folksong recordings from the *Onder de Groene Linde* collection that are available via the *Meertens Tune Collections* [2]. The goal of the study was to determine whether there is consistency in sung tonic pitch height across individuals when singing the same folksong independently of each other, across place and time. The results show that there is indeed some *inter-recording tonic pitch consistency* [3] in the recordings of a small collection of folksongs. This is the first study to suggest that Absolute Pitch Memory (APM) plays a role in oral transmission of folksongs.

Our working hypothesis is that all of the *tune families* [4] should show some level of inter-recording tonic pitch consistency as a sign that the melodies were memorized and transmitted on the basis of absolute pitch height, instead of just melodic contour (Dowling & Fujitani, 1971). If the empirical data supports this hypothesis, this would imply a role for APM in the oral transmission of folksongs, as suggested by earlier studies (e.g., Halpern, 1989; Levitin, 1994). The alternative hypothesis predicts that none of the investigated songs show inter-recording tonic pitch consistency, and instead the sung tonic

pitches can be expected to be uniformly distributed over recordings. Interpretations of these possible outcomes will be discussed later in this paper.

  We will first provide some background information on the topic of absolute pitch. After elaborating on the differences between two types of absolute pitch, AP vs. APM, we will present two lines of research on absolute pitch, followed by different theories on auditory memory, and an overview of the musical material used in this study. Subsequently, we will present our methods and results, followed by a discussion. Finally, we will present our conclusions and provide some suggestions for future research.

BACKGROUND

Traditional Absolute Pitch Versus Absolute Pitch Memory

Traditional Absolute Pitch (AP) – the ability to instantly identify and produce a certain tone without any reference note – is extremely rare; only 1 in 10000 individuals have this ability (Takeuchi & Hulse, 1993), and as a result it has been termed a gift by some researchers (e.g., Athos et al., 2007; Bachem, 1940; Gregersen, Kowalsky, Kohn, & Marvin, 2001). Others insist that anyone has the potential to acquire AP in the traditional sense, but that training in a critical period is needed to fully acquire the skill (e.g., Vitouch, 2003). Interestingly, all infants process pitch information in an absolute fashion. Some researchers have suggested that humans gradually shift to relative pitch processing when they get older and, as a result, lose their absolute pitch processing abilities (Saffran & Griepentrog, 2001). More recent evidence suggests that infants are, in fact, capable of relative pitch processing as well, if a slightly different task is used (Plantinga & Trainor, 2005). If infants possess both relative and absolute pitch processing capabilities and use them both depending on task requirements, why would adults lose either of the two?

  Perhaps adults do not lose either of the processing capabilities mentioned above. Over the past 25 years or so, researchers have partly shifted their attention to what some experts call "absolute tonality" (Vitouch, 2003) or "residual AP" (Deutsch, 2002). This has resulted in an increasing body of evidence in support of the notion that the capability to store absolute pitch information is in fact retained for melodies, even for adults with little formal musical training (Terhardt & Ward, 1982; Schellenberg & Trehub, 2003; Levitin, 1994; Vitouch & Gaugusch, 2000). Levitin (1994) has argued that AP possessors in the traditional sense – as opposed to the non-possessors – are able to label isolated pitches verbally, but that everyone possesses absolute memory for pitch (APM) in melodies. This claim – that APM is in fact widespread – is supported by evidence from two types of research. One of these lines of research uses identification tasks to address the issue, while the other uses production tasks. Some related work from both lines of research will be discussed in the next section.

Related Work on Absolute Pitch Memory

Identification tasks use an experimental design in which participants are presented with two versions of a musical excerpt. One of these versions is played to them in the correct key, while the other is transposed ("shifted") up or down by a certain number of semitones. The participants then have to judge which version is the correct one.

  An example of such an experiment was performed by Schellenberg and Trehub (2003). Their results indicate that even those participants with little formal musical training (and no reported AP capabilities) were able to identify the correct version of familiar TV program tunes well above chance. When the difference in tonic pitch between the correct and the incorrect version was one semitone, the participants chose the correct version 58% of the time. If there was a two-semitone difference between the excerpts, the participants even identified the correct version 70% of the time. These results show that at least some APM seems to be retained in adulthood.

  The second type of experiment investigates to what degree participants are able to produce or reproduce a melody in the correct key. Levitin (1994) showed that musically untrained participants achieved above chance pitch accuracy in a production task. In two trials, students with little formal musical training were asked to select a familiar song from CDs that Levitin provided. They were then asked to sing a self-selected part of this song at the correct pitch height (correct being: the pitch as it was recorded on the CD). About 40% of participants sang familiar rock tunes on the correct pitch in at least one of two trials,

suggesting that even singers with little to no formal musical training are able to produce or reproduce these pitches from memory to some extent. These results again indicate that APM is widespread.

More recently, Frieler et al. (2013) were able to replicate Levitin's experiment in six European labs, though with slightly smaller effect sizes than those reported in the original study. Findings varied widely between different laboratories, and as a result the authors stressed the importance of replication in music psychology research.

Additional evidence for APM using a production task design comes from within-subject experiments done by Bergeson and Trehub (2002) and Halpern (1989). In the first study, mothers were asked to sing a song to their infant that they would also sing in a non-experimental setting. They were instructed to sing the same song again a week later. Tonic pitch height was measured on both occasions. The mean tonic pitch deviation of the second, as compared to the first performance, was less than a semitone. In Halpern's experiment, participants had to sing the opening tones of holiday and childrens' songs in two trials. The opening tones of their second performance only deviated two semitones on average, compared to their first performance.

These last two studies by Bergeson and Trehub (2002) and Halpern (1989) are especially interesting for our study because the type of songs that the participants had to sing closely resemble the type of songs used in the current study in that there was no standardized version available to the participants. In other words, participants did not have an external reference such as sheet music or a recording. Instead, the absolute pitch information was retrieved from memory only, and therefore this memory of earlier performances, including the participants' performance in the first trial, served as an internal reference for their second performance.

Song Memory

How do singers recall melodies from long-term memory? Some researchers have proposed that attributes such as pitch, tempo, and timbre are stored in a multiple-trace memory system (Levitin & Rogers, 2005) that is connected to a perceptual system (Dalla Bella, Peretz, & Aronoff, 2003). Levitin and Rogers suggest that memory for AP is a low-level feature that co-exists with an abstract memory for other features such as the sequence of relative pitches in the melody and, most likely, the emotion associated with the piece as well (Eschrich, Münte & Altenmüller, 2008).

Furthermore, it has been suggested that imagery plays a role in the retrieval of a memory. Before and during performance of a melody, one actively imagines the melody. When trying to recall and reproduce it, one combines all features, such as absolute and relative pitch, tempo, and timbre into the end product, the melody itself, while receiving immediate perceptual feedback from one's own voice. The more often a melody has been rehearsed, the more accurate the representations become (Keller, Cowann, & Saults, 1995). This in turn results in better recall, and thus reproduction of the melody. This could be a reason why musicians often perform better in song (re)production experiments. All of this is consistent with neurological evidence from Zatorre and Halpern (1993), demonstrating that once auditory cortical areas have been damaged, for example due to a focal auditory cortex lesion, perceptual and imagery deficits co-exist, resulting in worse retrieval of memories that include auditory features.

Finally, procedural memory has been suggested to be involved in the retrieval and (re)production of a melody representation (Brown, Ngan, & Liotti, 2007). Located in the motor cortex, a specific area representing the vocal tract could potentially be connected to the several representations of the melodies stored in memory as well as the vocal tract itself. In this scenario, it is as though the vocal tract "remembers" the exact tension associated with each pitch in the representation. However, this possibility remains mostly theoretical, as little evidence of such procedural memory has been presented in the vocal domain thus far(see, e.g., Brown & Palmer, 2012). Also, as Levitin (2013) has recently noted, procedural memory cannot solely account for the accurate pitch reproduction that adults are capable of.

Materials: folksongs

In this study, all recordings were part of the "Onder de Groene Linde" (OGL) collection as recorded by Will Scheepers and Ate Doornbosch in the 1950s and onwards, available online via the Meertens Tune Collections. These recordings feature mostly older adults with little to no formal musical training, singing

Dutch folk songs from memory only. Like in some of the production studies mentioned earlier (e.g., Bergeson & Trehub, 2002), there was no standardized version available for these songs.

Experts have grouped the recordings in the Meertens Tune Collections into tune families of closely related variants, which enabled us to compare the tonic pitch of different variants to each other, and thereby explore the potential role of APM in the memory of songs transmitted in oral traditions. The oral transmission of folksongs is highly influenced by memory, as well as perception, performance, and creativity (van Kranenburg, 2012). It is therefore important to show that there is at least some pitch consistency on a between-subject scale. Once one is able to define the components of auditory memory that play a central role in oral transmission, one can study these components and use the resulting knowledge for a theory of oral transmission of folksongs.

Analyses of two partly different datasets were performed to investigate the role of APM in folksongs. Dataset A was used for a between-tune family analysis examining the inter-recording tonic pitch consistency of variants from seven tune families. Dataset B consisted of a larger set of recordings of two tune families, allowing for a within-tune family analysis of subsets grouped according to three factors that might influence the songs' inter-recording tonic pitch consistency.

The first factor that could possibly influence our results is the lyrics of the song. Different textual versions of one melody may belong to the same tune family. Contextual cues and affect (Bergeson & Trehub, 2002) have been proposed to facilitate retrieval of performance details of musical melodies. One of these contextual cues could be the lyrics accompanying the melody (cf. Levitin & Rogers, 2005). Secondly, gender may influence the pitch on which a folksong is usually sung, because of the typical ranges of men's and women's voices (Titze, 1989). Finally, if there is a role for APM in oral transmission of folksongs, the tonic pitch of folksongs might develop independently in different geographical regions, predicting possible different mean tonic pitches in recordings from different regions. In the ensuing section, the methods and the results of the between-tune family analysis will be described.

DATASET A: BETWEEN TUNE FAMILY ANALYSIS

Method

To investigate the role of APM in folksongs, we analyzed the tonic pitches sung in the recordings available via the Meertens Tune Collections post hoc, using comparable methods to those used in the production studies mentioned earlier in this paper (e.g., Halpern, 1989; Bergeson & Trehub, 2003). The recordings in the OGL databank were of various lengths, ranging from one to ten minutes. We decided to compare the first verses of each recording, as other factors such as pitch height fluctuations (i.e. drift) might have intervened if first verses had been compared with later verses. Moreover, there is evidence that the first verse best reflects the representation of the tonic pitch retrieved from memory (e.g., Halpern, 1989; Klinger, Campbell, & Goolsby, 1998; Smith, 1991).

In our between-tune family analysis, we examined sung tonic pitches in recordings of 7 tune families. The tune families selected for the between-tune family analysis were: "Daar was laatst een meisje loos" (N=20), "Er reed er eens een ruiter" (N=20), "Het was laatst op een zomerdag" (N=20), "Al is ons prinsje nog zo klein" (N=19), "Het vrouwtje van Stavoren" (N=20), "Mijn vader zei laatst tegen mij" (N=20), "Wat hoor ik hier in het midden van de nacht" (N=20). The particular tune families used were chosen based on the availability of at least 20 recordings of these tunes [5]. One recording of "Al was ons prinsje nog zo klein" was removed from the data set, because it contained an instrument playing the tune rather than someone singing it, leaving 19 recordings of this tune family for data analysis.

Yin, a pitch detection algorithm (de Cheveigné & Kawahara, 2002), was used to detect the fundamental frequencies of various pitches sung in the recordings. From the output of this algorithm, a density estimation of the distribution of pitches was computed using a Gaussian Kernel following the method of Biró, van Kranenburg, Ness, Tzanetakis, & Volk (2012). The first author then checked the frequencies achieving the highest density with the help of a tone generator [6], to determine which of these frequencies represented the tonic frequency.

Quantitative analysis with circular statistics

Pitch can be represented in several ways. We will adopt the representation proposed by Shepard (1982), who distinguished between pitch height and pitch chroma. The latter is a common grouping principle, also referred to as pitch class, where a tone of, e.g., 440 Hz (orchestral A4) is considered the same as 220 Hz (A3) or 880Hz (A5). Since there is also some evidence that the brain represents pitch along these two dimensions (Warren et al., 2003), we used pitch chroma, or octave normalization, as a way to group the pitches analyzed.

Because of the circular nature of octave-normalized pitch (Levitin, 1994), a circular test such as the Rayleigh test was needed to determine or reject uniformity of the data (Fisher, 1993). We ensured that the investigated pitch distributions fit a unimodal distribution through Hartigan's dip test (Hartigan & Hartigan, 1985) before performing the Rayleigh test, which presupposes unimodal distribution. If Rayleigh's test returns a result equal to or below .05, the frequency of occurrence of the sung pitches significantly deviates from a uniform distribution, in favor of a unimodal distribution.

To convert our measurements (fundamental frequency of the tonic) to a scale suitable for circular statistics, we first converted our frequencies to MIDI note numbers, an equivalent of pitch on a linear scale. Because the octave can be divided in twelve semitones, or in this case twelve MIDI note numbers, every MIDI note number can be thought of as 30 degrees. The MIDI note numbers were then converted to angles (in degrees) by multiplying them by 30, ensuring successful octave generalization for our obtained results [7]. To do this, we used Oriana [8], a tool developed specifically for this type of statistics.

Baseline

To determine how much of the tonic pitch consistency might be caused by the vocal range of the singers, and the melodic range of the folk songs, we ran a Monte Carlo simulation of each tune family. The vocal ranges used for this simulation were based on Moore, Staum and Broton's (1992) analysis of the singing ranges of geriatric persons between 60 and 110 years of age, as such roughly matching the singers recorded in the Meertens database.

The average vocal range for male singers was set at G#2 to D4 and the average female range was set at F3 to C5. As the female to male ratio in our dataset was 4:1, we gave higher probability (.8) to a female vocal range. The vocal range was represented in MIDI note numbers, on a continuous scale.

The melodic ranges of the simulated songs were sampled from the available transcriptions for the tune families used in this study. The pitches of the first five notes of each transcription were examined to determine the highest and lowest starting pitch. The melodic range was represented in semitone steps between the highest and lowest starting pitch.

The simulation of the tonic pitch departed from the lowest starting pitch, such that the melodic range fit into the vocal range. For instance, if the vocal range sampled was that of a female singer, 53.0 to 72.0 in continuous MIDI note numbers, and the melodic range was seven semitones, the lowest starting pitch would be sampled from values between 53.0 and 65.0. Finally, the position of the nearest tonic pitch above the lowest starting pitch – information that was also derived from the transcriptions – was determined, providing the sampled tonic pitch.

We ran 100 simulations of 20 tonic pitches and, for each simulation, we performed Hartigan's dip test. If the test implied a non-unimodal distribution, we excluded the simulation. Otherwise, we performed the Rayleigh test and collected the resulting p-values. The mean p-value of the Rayleigh tests on the Monte Carlo simulations indicates whether there is a bias for a unimodal distribution based on the vocal range of the singers and melodic range of the songs.

Results

The results of the between tune family analysis are shown in Figure 1. The null hypothesis of bimodal or multimodal distribution could be rejected for all tune families (p>0.6 throughout). The null hypothesis of uniformity was rejected for the tune family "Daar was laatst een meisje loos" ("**Loos**"), $r = .49$, $p < .01$. The null hypothesis of uniformity could not be rejected however for "Er reed er eens een ruiter" ("**Ruiter**"), $r = .35$, $p = .08$, "Het was laatst op een zomerdag" ("**Zomerdag**"), $r = .16$, $p = .61$, "Het vrouwtje van Stavoren" ("**Stavoren**"), $r = .19$, $p = .51$, "Al was ons prinsje nog zo klein" ("**Prinsje**"), $r = .24$, $p = .34$,

"Mijn vader zei laatst tegen mij" ("**Vader**"), *r* =.28, *p* =.21 and "Wat hoor ik hier in het midden van de nacht" ("**Nacht**"), *r* =.12, *p* =.74.
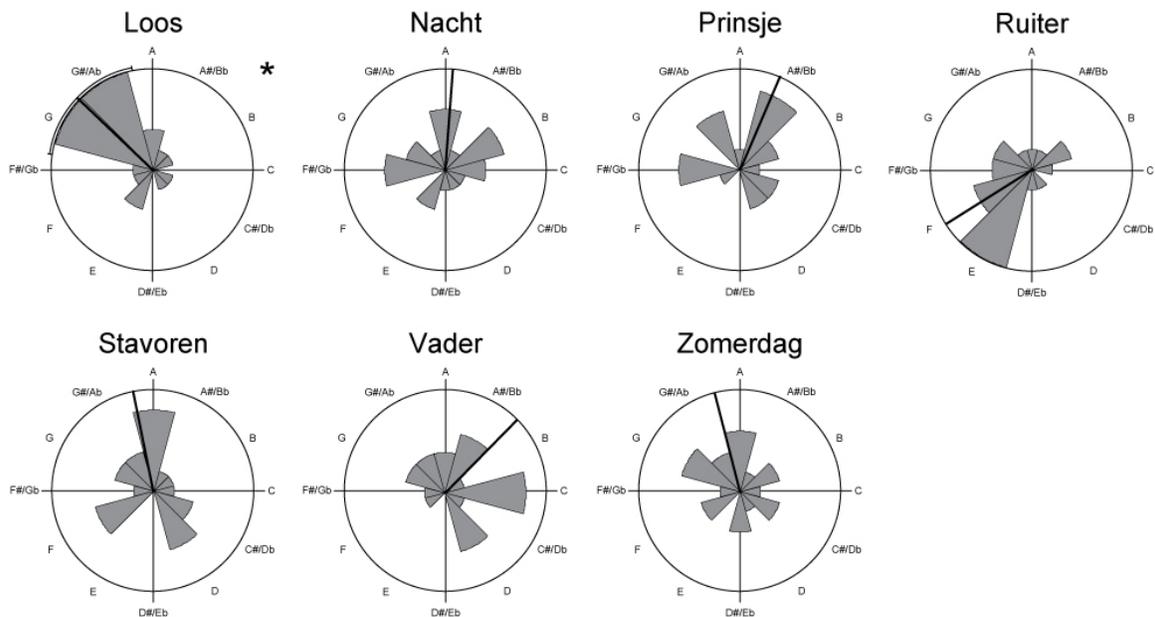

Fig. 1. Between-tune family results. *Circular histograms representing the distribution of pitch chroma for seven tune families. Arrows indicate the mean vector. Significant results are marked with an asterisk and confidence interval. The legend shows how pitch chroma are mapped onto degrees in the diagram.*

The baseline simulation of the vocal and melodic range yielded an average p-value of p=0.39 on the Rayleigh tests. Three simulations were excluded due to a significant result from Hartigan's dip test. This implies that the evidence for unimodal distribution is not likely to be caused by the restrictions of the singers' vocal range and the song's melodic range.

Therefore, the significant results of **Loos** indicate that there may be some APM involved in the recall and reproduction of this particular folksong. The remaining tune families may not have yielded significant results for various reasons: for example, the dataset simply might not be large enough to observe significant deviations from the uniform distribution because of possible factors of variance such as lyrics, gender, and geographical origin involved. To test these assumptions, two of the tune families of which there were more recordings available were again used in our within-tune family analysis.

DATASET B: BETWEEN & WITHIN TUNE FAMILY ANALYSIS

Method

Dataset B consisted of two tune families, "Wat hoor ik hier in het midden van de nacht" (N=53) and "Mijn vader zei laatst tegen mij" (N=67). The procedure used to determine the fundamental frequencies of the first verse of these recordings was similar to the one used for dataset A. This time, gender of the singer(s), geographical origin, and difference in lyrics of each recording were noted for additional analysis.

Baseline

The full tune families as well as the subsets of female and male singers, geographical origins, and lyrics were also subjected to a simulation of starting pitches, corresponding to the simulations for dataset A. The sample sizes in the simulations were set to correspond to the sizes of the investigated datasets.

Results

The results of the between- and within-tune family analysis performed on dataset B are shown in Figure 2, this time comparing a greater number of variants than reported in Figure 1. Following a Hartigan's dip test, the null hypothesis of bi- or multimodality was rejected for all subsets of the tune families with the exception of the geographic subset of the province Drenthe, for which Hartigan's dip test yielded D = 0.1191 and p < 0.03. In consequence, we performed Rao's spacing test (Jammalamadaka & SenGupta, 2001) instead of the Rayleigh test to investigate the distribution.

  The null hypothesis of uniformity was rejected for "Mijn vader zei laatst tegen mij" ("**Vader**") (N=67)*, r =*.26*, p* <.01. It could not be rejected for "Wat hoor ik hier in het midden van de nacht" ("**Nacht**") (N=53), *r =*.35, *p* =.06. In the simulation of starting pitches of **Vader**, the Rayleigh tests achieved an average p-value of p=0.36, indicating that there is no tendency towards unimodal distribution based on vocal range and melodic range alone.

  Next, additional analyses were done for "Mijn vader zei laatst tegen mij" to determine whether there was any effect of the lyrics that were sung. There were two major textual versions among the recordings, **Text Vader** (N=33) and **Text Boerenzoons** (N=32). A Rayleigh test was conducted to determine uniformity. The null hypothesis of uniformity was rejected for both **Text Vader**, *r =*.36, *p* <.02, and **Text Boerenzoons**, *r =*.34, *p* <.02. Interestingly, the means (17 and 111 degrees, respectively, see Figure 2) for both textual versions were as much as 3 semitones apart, even though range and contour of both versions were highly similar. The simulation of vocal range did not show a bias for unimodal distribution, with the mean p-value at p=.47 for **Text Vader,** and at p=.55 for **Text Boerenzoons.** For both subsets, the dip test indicated unimodality for all simulations.

  Gender may also be a factor of variance in the sung pitches, as discussed in the introduction. Based on the Rayleigh test, the null hypothesis of uniformity was rejected for recordings of *Vader* that were sung by female singers, *r =* .29, *p* < .02 (N=49) but not for recordings the same song that were sung by male singers, *r =*.24, *p* =.49 (N=12). Similarly, the null hypothesis of uniformity was rejected for recordings of **Nacht**, that were sung by female singers, *r =*.39, *p* < .01 (N=33) but not for recordings of **Nacht** that were sung by male singers, *r =*.20, *p* =.56 (N=14). This implies that the pitch consistency in female recordings is remarkably higher than the consistency in male recordings. However, there were relatively few male singers, which could have contributed to the non-significant results. The baseline simulations of the male and female subgroups showed no tendency towards significant p-values on the Rayleigh tests, with means of p=0.36 for the female subgroup of **Vader** (2 excluded simulations due to significant dip test)*,* and p=0.15 for the female subgroup of **Nacht** (no excluded simulations).

  The last factor that we included in our analysis was geographical origin. In the Meertens Tune Collections, the origin of the recordings is included in the description. There are 12 provinces in the Netherlands (see Figure 3), most of which have their very own identity and culture. Most provinces did not have enough data points to do a statistical analysis based on geographical origin, even in our within-tune family dataset. Two provinces, however, did have a reasonable number of recordings to look for a trend in the data. This was the case for particular tune families in the second dataset: Groningen for "Mijn vader zei laatst tegen mij" (N=22) and Drenthe for "Wat hoor ik hier in het midden van de nacht" (N=16) (see Figure 3). A Rayleigh test was conducted for recordings from both provinces to determine uniformity when geographical origin was included as a factor. The null hypothesis of uniformity was rejected for recordings of "Mijn vader zei laatst tegen mij" originating from Groningen, *r =*.38, *p* < .05, but could not be rejected for recordings of "Wat hoor ik hier in het midden van de nacht" from Drenthe, *p* > .1 (see Figure 2). For both cases, the simulations of the geographical subsets resulted in mostly non-significant p-values. The mean p-value for the Rayleigh tests on the Groningen subset for **Vader** was p=0.5, and for the Drenthe subset of *Nacht* it was p=.62. No simulations were excluded for the geographical subsets because of non-unimodal distributions. This supports our assumption that higher pitch consistencies are caused by geographical factors, rather than by the melodic ranges of the songs.
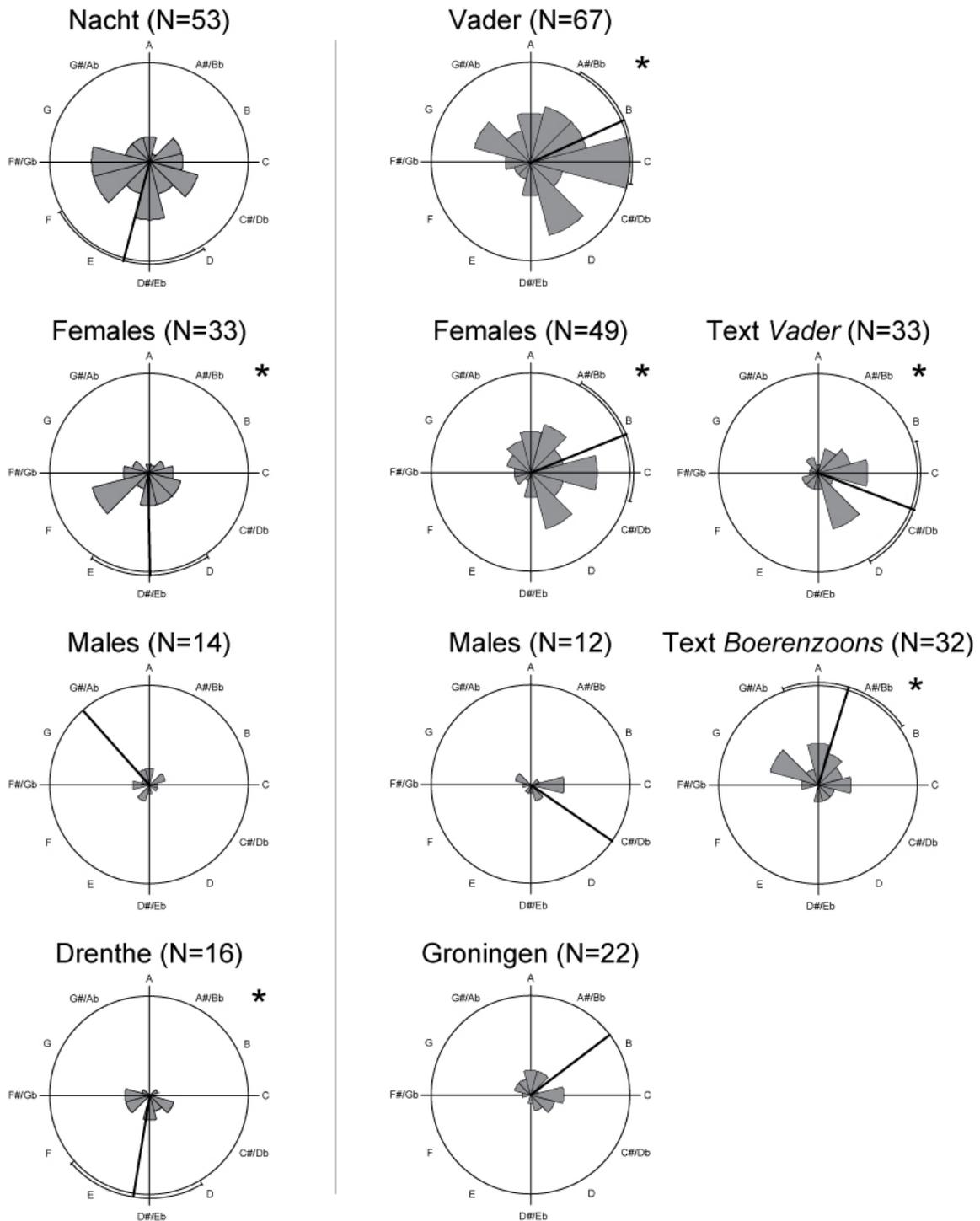
Fig. 2. Within tune family results. *Circular histograms representing the distribution of pitch chroma for the tune families Nacht (left panel) and Vader (right panel). Arrows indicate the mean vector. Significant results are marked with an asterisk and confidence interval. The legend shows how pitch chroma are mapped onto degrees in the diagram. N.B. Subsets of textual versions and gender do not add up to the complete dataset because there were some recordings with other texts, and unspecified gender of the singers.*
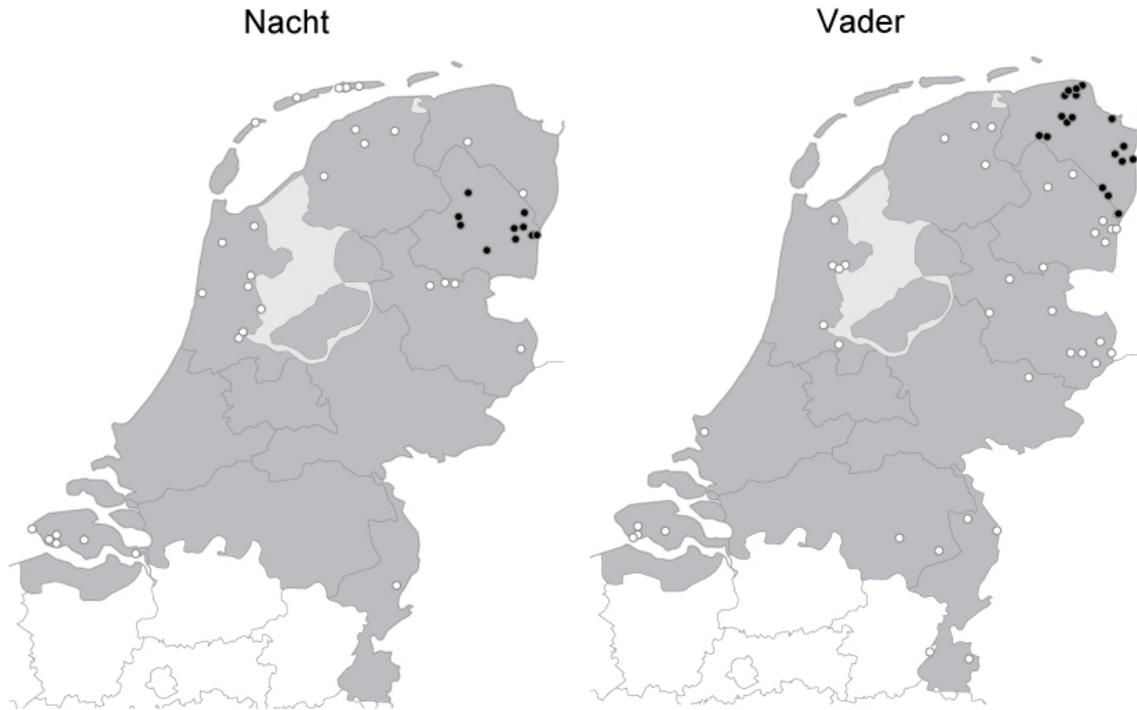
Fig. 3. *Geographical distribution of the recordings used for Nacht (left panel) and Vader (right panel) over the Netherlands. Black dots mark recordings in a province that have a reasonable number of recordings for analysis (i.e. Drenthe for Nacht and Groningen for Vader), white dots show the remaining recording locations from the respective tune families. N.B. A single dot can represent multiple recordings.*

DISCUSSION

A Role for Absolute Pitch Memory in Oral Transmission of Folk Songs

The results indicate that absolute pitch memory plays a role in the oral transmission of folksongs. As there was no standardized version of these folksongs available to the singers, these singers reproduced the songs solely from memory. However, not all tune families showed inter-recording tonic pitch consistency. In this discussion, several potential explanations for these mixed results are addressed.

If there is no standard version, one might expect that subjects sing the tune on a different tonic pitch every time, and therefore the absolute pitch information is not reinforced and stored as part of the mental representation of the song. In that case, a standardized version of the song might be needed as a reference to establish absolute pitch information in memory, as Levitin (1994) already hypothesized. This reference could be the original version, played by musical instruments, or a written standard version of the song.

However, a standard version often does not exist for folksongs, especially those in the *Onder de Groene Linde* collection because its collectors, Will Scheepers and Ate Doornbosch, had a preference for folksongs which were not explicitly learned (for instance, songs learned at school or church) at school or from church song books. "Daar was laatst een meisje loos" might be an exception, as it did appear in a standardized fashion in songbooks from the early 20th century onwards. Interestingly, this was the only song showing significant inter-recording tonic pitch consistency in between tune family analysis. However, this does not explain why the two larger tune families, both of which did not have a standardized version, showed inter-recording tonic pitch consistency after controlling for certain variables such as gender, which we will discuss later.

Another explanation for the mixed results could be that the singers were not able to produce the tone they wanted to produce (Levitin, 1994). In that case, the mental representation of the melody might have been correct, but the singers were unsuccessful in reproducing the tone they had in mind. These

unsuccessful attempts could be due to physical tone production problems (Levitin, 1994) and/or a lack of musical training. In general, pitch memory and the perception of pitch relations are a function of musical experience (Krumhansl, 2000; Schellenberg & Trehub, 2003). The singers in the recordings of the *Onder de Groene Linde* collection had little to no formal musical training and therefore might have been less precise in reproducing the tone they had in mind.

Through our baseline study we show that the range of the starting pitches of a melody do not seem to constrain the choice of pitch height significantly. However, one aspect to investigate in the future is whether the contour of the starting pitches has an influence on the chosen pitch height. For example, a jump of a fifth following the starting pitch might constrain the singer's choices in a different way than an ascending scale spanning a fifth.

There are several other reasons for a possible lack of pitch consistency in the recordings related to the particular materials used in this study. For example, in most of their recording sessions, Doornbosch and Scheepers asked the singers to sing more than one folksong. If there is no pitch consistency to be found, this might be due to influences of performances of other songs during the same recording session. According to Levitin (1994), the tonal center of the first song that was sung during a session might simply determine the tonal center of the second song as well.

We cannot compare our results to Levitin's (1994) and Frieler et al.'s (2013) results in terms of effect sizes. Frieler and colleagues estimated the power and effect sizes of their results by testing the number of 'hits' in their trials by means of a Goodness-of-Fit- χ^2-test. However, due to the fact that there are no standardized versions available of the folk songs used in our study, we do not have 'hits', and our analysis relies purely on the distribution of tonic pitches.

Gender, Lyrics and Geographical Origins

In the introduction, we suggested three factors that might influence the results. We have tested for these three factors in our second study, but the results are tentative. To truly control for these variables, one would require a larger dataset than the one available in this study. In a larger dataset, if one would control for the factors of gender, geography, and contextual factors, according to the "APM is widespread" hypothesis one might find inter-recording tonic pitch consistency in every subset of variants from the same geographic region, with the same lyrics, or sung by singers with the same gender.

Different means in tonic pitch for women and men for the same song were obtained, perhaps due to the different fundamental frequencies men and women prefer for speech and song. These differences can in turn be attributed to physiological differences, for example the size of the larynx (Titze, 1989). There was a remarkably higher inter-recording tonic pitch consistency between female recordings for both songs in the second dataset. This may have been due to puberty voice changes for male subjects (Harries, Walker, Williams, Hawkins, & Hughes, 1997). As these songs are often learned at young age, male singers might have learned these songs before puberty. The voice changes associated with male puberty might have distorted their reproduction of the melody as they had learned it. It was thus harder to reproduce the initial learned pitches. However, this remains mostly speculative, because relatively few recordings of male singers were available.

Regional variance may also explain the lack of pitch consistency in the smaller datasets. If geographical origin is indeed a factor of pitch variance, this might lead to disparities in the tonic pitch, especially for the smaller tune families, because the song might have been sung on a different tonic pitch in different parts of the Netherlands (see Figure 3). Those provinces that did have a reasonable number of recordings of particular tune families for statistical analysis showed inter-recording tonic pitch consistency, indicating that the same songs may be sung on different tonic pitches in different geographical areas. However, a bigger sample is needed to draw more definite conclusions about this matter.

Interestingly, the means for two textual versions of one of the tune families in the second dataset were as much as 3 semitones apart, suggesting that tune families with different lyrics but highly similar relative melodies may have two separate absolute pitch representations in memory. This is in accordance with the proposals by Bergeson and Trehub (2002) and Levitin and Rogers (2005), stating that lyrics accompanying the melody can facilitate retrieval of performance details of musical melodies. However, a larger dataset is needed to be able to test these interpretations.

CONCLUSIONS

In the current study, we tested whether there was any inter-recording tonic pitch consistency in recordings of folksongs available via in the Dutch Song database. In dataset A, one tune family, "Daar was laatst een meisje loos", showed significant inter-recording tonic pitch consistency based on a collection of twenty recordings. In dataset B, one tune family, "Mijn vader zei laatst tegen mij", showed tonic pitch consistency over the whole dataset, but also when grouped based on lyrics, suggesting a possible role for contextual information such as the lyrics. The second tune family in this dataset, "Wat hoor ik hier in het midden van de nacht" showed near significant inter-recording tonic pitch consistency. When we controlled for gender, both these tune families showed tonic pitch consistency for female, but not for male singers. Also, for both these tune families there was significant consistency for tonic pitch when grouped by geographical origin of the recordings.

Nevertheless, to really identify the influence of these factors a larger dataset is needed. For this, the Meertens Tune Collections continues to be an excellent source. The recordings in the database are well documented and therefore easily retrievable. They have been grouped by tune family, which is important when looking at phenomena related to oral transmission of folksongs.

Further research in this direction would be a valuable addition to our knowledge on the role of absolute pitch memory in oral transmission. Such research could be performed on recordings from other song collections, using similar methods to those used in this study.

ACKNOWLEDGEMENTS

NOTES

[1] Correspondence can be addressed to: Berit Janssen, University of Amsterdam, mail: berit.janssen@meertens.knaw.nl

[2] http://www.liederenbank.nl/mtc/

[3] *Inter-recording tonic pitch consistency* is consistency in sung tonic pitch height (or pitch chroma) across individuals when these individuals sing the same folksong independently of each other.

[4] The concept of *tune family* was developed by Samuel Bayard (1950). He defined it as: "a group of melodies showing basic interrelation by means of constant melodic correspondence, and presumably owing their mutual likeness to descent from a single air that has assumed multiple forms through processes of variation, imitation, and assimilation" (Bayard, 1950, p. 33).

[5] Some of these *tune families* had more than twenty recordings available. However, the first twenty recordings were always selected, based on their number in the database. The tune families investigated in this study have the following identifiers in the Meertens Tune Collections: 1497_1, 9744_1, 9779_0, 288_1, 3676_1, 5899_0 and 9107_0.

[6] http://www.audionotch.com/app/tune/

[7] The complete calculation used to convert our measured tonic frequencies to degrees was
($n = 12 * (\log(f/220) / \log(2)) + 57) *30$, where n = MIDI tones and f = measured tonic frequency in Hertz.

[8] Windows version 4.01, Kovach Computing Services, Pentraeth, Anglesey, Wales, U.K.

REFERENCES

Athos, E. A., Levison, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N., & Gitschier, J. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences,* USA, *104*, 14795-14800.

Bachem, A. (1940). The genesis of absolute pitch. *Journal of the Acoustical Society of America*, *11*, 434-439.

Bayard, S. (1950). Prologomena to a study of the principal melodic families of British-American folk song. *Journal of American Folklore*, *63*(247), 1-44.

Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mother's songs to infants. *Psychological Science*, *13*, 72-75.

Biró, D. P., Kranenburg, P. van., Ness, S. R., Tzanetakis, G.. & Volk, A. (2012). Stability and variation in cadence formulas in oral and semi-oral chant traditions – A computational approach. *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, 98-105.

Brown, R. M., & Palmer, C. (2012). Auditory- motor learning influences auditory memory for music. *Memory & Cognition*, *40*, 567-578.

Brown, S., Ngan, E., & Liotti, L. (2007). A larynx area in the human motor cortex. *Cerebral Cortex*, *18*, 837-845.

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics*, *65*, 1019-1028.

De Cheveigné, A., & Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music, *The Journal of the Acoustical Society of America*, *111*(4), 1917-1930.

Deutsch, D. (2002). The puzzle of absolute pitch. *Current Directions in Psychological Science*, *11*, 200-204.

Dowling W. J. & Fujitani D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, *49*, 524–31

Eschrich, S., Münte, T. F., & Altenmüller, E. O. (2008). Unforgettable film music: the role of emotion of episodic long-term memory in music. *BMC Neuroscience*, *48*(9).

Fisher, N. I. (1993). *Statistical analysis of circular data.* Cambridge: Cambridge University Press.

Frieler, K., Fischinger, T., Schlemmer, K., Lothwesen, K., Jakubowski, K., & Müllensiefen, D. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae*, *17*(3), 334-349.

Gregersen, P. K., Kowalsky, E.,Kohn, N., & Marvin, E. W. (2001). Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment. *American Journal of Medical Genetics*, *98*, 280-282.

Grijp, L. P. (2008). Introduction. In L. P. Grijp & I. van Beersum (Eds.), *Under de Green Linden: 163 Dutch ballads from the oral tradition.* (18-27). Music & Words.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory and Cognition*, *17*, 572-581.

Harries, M. L. L., Walker, J. M., Williams, D. M., Hawkins, S., & Hughes, I. A. (1997). Changes in the male voice at puberty. *Archives of Disease in Childhood*, *77*, 445-447.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *Annals of Statistics*, *13*(1), 70–84.

Jammalamadaka, S. R., & SenGupta, A. (2001). *Topics in circular statistics*. Singapore: World Scientific Press.

Keller, T. A., Cowann, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21*(3), 635-645.

Klinger, R., Campbell, P. S., & Goolsby, T. (1998). Approaches to children's song acquisition: Immersion and phrase-by-phrase. *Journal of Research in Music Education*, *46*(1), 24-34.

Kovach, W. L. 1994. Oriana. Kovach computing services. Anglesey, Wales, United Kingdom.

Krumshansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*, 159-179.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence for the production of learned melodies. *Perception & Psychophysics*, *56*, 414-423.

Levitin, D. J., &Rogers, S. E. (2005). Absolute pitch: Perception, coding and controversies. *Trends in Cognitive Science*, *9*, 26-33.

Levitin, D. J. (2013). Commentary on "Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae*, *17*(3), 350-355.

Moore, R. S., Staum, M. J., & Brotons, M. (1992). Music preferences of the elderly: Repertoire, vocal ranges, tempos, and accompaniments for singing. *Journal of Music Therapy*, *29*(4), 236-252.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*, 1-11.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology,*

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*, 262-266.

Shepard, R. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, *89*(4), 305–333.

Smith, D. S. (1991). A comparison of group performance and song familiarity on cued recall tasks with older adults. *The Journal of Music Therapy*, *28*(1), 2-13.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*, 345-361.

Terhardt, E, & Ward, W. D. (1982). Recognition of musical key: Exploratory study. *The Journal of the Acoustical Society of America*, *72*(1), 26-33.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*(4), 1699-1707.

van Kranenburg, P. (2012). *A computational approach to content-based retrieval of folk song melodies*. (Unpublished doctoral dissertation). Utrecht University, Utrecht.

Vitouch, O. (2003). Absolutist models of absolute pitch are absolutely misleading. *Music Perception*, *21*, 111-117.

Vitouch, O., & Gaugusch, A. (2000). Absolute recognition of musical keys in non-absolute-pitch-possessors. In C. Woods, G. Luck, R. Brochard, F. Seddon, J. A. Sloboda (Eds.), *Proceedings of the 6ᵗʰ International Conference on Music Perception and Cognition.* Keele, UK: Department of Psychology, Keele University.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(17), 10038–42.

Zatorre, R. J., & Halpern, A. R. (1993). Effects of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia*, *31*(3), 221-232.