Absolute Pitch in Oral Transmission
of Folk Tunes
as Constrained Random Walks

KLAUS FRIELER
University of Music FRANZ LISZT Weimar

ABSTRACT: In this commentary, I would like to add a few of our own, still unpublished, empirical observations concerning the possible role of absolute pitch memory (APM) in the oral transmission of folksongs. This empirical data poses some questions on the likelihood of the observed inter-recording tonic pitch consistency of Olthof, Janssen & Honing (2015) and how these could come about. Based on simulations of absolute pitch class of tonics during oral transmission of folk songs, I argue that the interplay of melodic range and vocal range might actually be the main reason for the observed non-uniformity, in contrast to the conclusions presented in Olthof et al. (2015). However, this does not invalidate the therein presented evidence, but makes the case more puzzling, consequently calling for more empirical research on the interaction of melodic and vocal range and latent APM as well as for more detailed modeling of oral transmission of folk songs.

IN their very inspiring article Olthof, Janssen & Honing (2015) study the role of absolute pitch memory (APM) in the context of oral folk song transmission. By investigating original recordings of several tune-families from the Meertens Tune Collection the authors found—at least in some cases—a deviation from uniformity in the distribution of the tonic pitches in the tune family, suggesting that not only musical syntactical information and lyrics are transmitted from singer to singer, but also the absolute position of a melody in pitch space. The authors also found some interesting differences with respect to gender, lyrics, and geographical origin of the recording. These results provide further evidence for the fact that latent AP is widespread (Levitin, 1994; Frieler et al., 2013).

In the following, I will present some hitherto unpublished experimental data directly related to the topic. First, in a study on drift and intonation in unaccompanied solo singing, Mauch, Frieler & Dixon (2013) allowed participants to sing several renditions of "Happy Birthday". The absolute tonic pitches of the renditions were measured but were not presented in the final publication, because this data was only marginally related to the primary research question. Second, during the International Summer School for Systematic Musicology 2011 in Jyväskylä, Finland (ISSSM2011), a group of students under my supervision set up two artificial oral transmission chains to study the process *in situ*. This study provides direct insight into how the absolute pitches of songs change during face-to-face learning processes. Finally, simulations of absolute tonic pitch classes based on the gathered evidence will be presented and discussed.

FINDINGS OF OLTHOF ET AL. (2015)

Olthof et al. (2015) conducted two studies using recordings of Dutch folk songs from the Meertens Tune Collection, more specifically the "Onder the groene Linde" collection, which comprises a large set of recordings of unaccompanied folk songs by singers without formal training that were collected from 1950 onwards from various locations across the Netherlands. The songs in the Meertens Tune Collection are grouped into tune families, which are defined by high melodic similarity even though the lyrics can differ

completely. Grouping into tune families was done by expert music ethnologists. In their first study, the authors examined 20 songs each from 7 different tunes families. They used only the first verses of each song to circumvent modulation and pitch drift issues and determined the absolute pitch height of the tonic using signal processing and manual evaluation. Using circular statistics, they found a significant deviation from uniformity for two of the seven tune families (by first ruling out bimodality using Hartigan's dip test), which indicates the prevalence of certain tonics, and hence, according to their hypothesis, the influence of latent APM in oral transmission. Furthermore, they calculated a baseline distribution of tonic pitches, based on a melodic range of the first five notes of each song and using average melodic ranges for elderly female and male singers taken from the literature (Moore, Staum and Broton, 1992). This is an important point, as will be demonstrated below, since it seems that using only the first five notes of a song for determining the melodic range as opposed to using the whole tune makes a significant difference. They simulated 100 lowest starting pitches and derived from this the pitch classes of the tonics with the help of available transcriptions of the songs. (However, it is not entirely clear from the text, if they performed this for each song independently or only for 20 songs in total.) Furthermore, they found an average p-value of 0.39 for the Rayleigh tests of circular uniformity (after excluding apparently multi-modal simulations according to Hartigan's dip test), thus concluding that the interaction of melodic and vocal range does not explain the observed deviations from uniformity in the data. However, reporting only the average p-values seems not to be the best possible choice. Particularly, it would have been more interesting to know the proportion of significant Rayleigh tests in relation to the expected number of significant results, e.g., by reporting a Bayes factor of expected to observed significances.

In their second study, Olthof et al (2015) focused on two tune families with a larger number of songs (N=52 and N=67), which were also differentiated with respect to lyrics, geographical origin and gender of the singer. Using the same method as in the first study, they found significant deviations from uniformity for one of the tune families (VADER) but not for the other (NACHT). Moreover, they found significant deviations from uniformity in the VADER tune family for two different text variants. For both tune families, the female version deviated from uniformity, but not the male versions. Finally, geographical subsets from the Drenthe and Groningen regions were analyzed for both tune families. One tune family deviated from uniformity (the Drenthe subset of the NACHT family). All baseline simulations yielded non-significant average p-values, but the same criticism to using this statistic as mentioned above applies here as well.

UNACCOMPANIED SOLO SINGING

In the first part, I will report on some evidence from a study on intonation in unaccompanied solo singing which is basically supporting some of the result by Olthof et al. (2015). The results presented here will also serve as input for a model of transmission of the tonic in the third part of my commentary.

Method

The main objective of Mauch et al. (2013) was to study the influence of different auditory feedback conditions on intonation and drift in unaccompanied solo singing. Twenty-eight participants with a broad range of singing skills and musical experiences were asked to sing three consecutive runs of "Happy Birthday" under different feedback conditions. The first condition was a control condition without
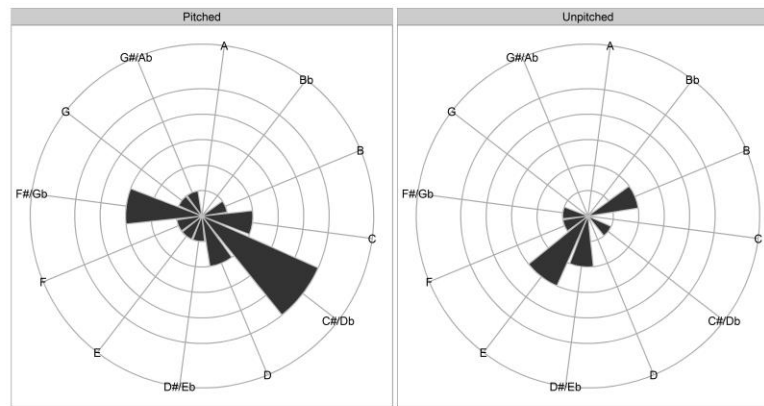
Fig. 1. Circular histograms of chosen tonics in the experiments of Mauch et al. (2013), using rounded median values of tonic pitches across conditions mapped. Left panel: Pitched click track (N=18), right panel: unpitched click track (N=10).

any manipulation. In the second condition, participants received a masking noise via headphones during the second run, and in the third condition they were instructed to only imagine but not to sing aloud the second run. All renditions were recorded and F0 annotations were extracted using the pYIN algorithm [1].

Results

As hypothesized, significant overall drifts of up to 60 cents could (occasionally) be observed. Moreover, by investigating pitch and interval error distributions, the authors could show that the data is best described by an intonation model in which pitches are performed as intervals with respect to a tune-local tonic compared to models of pure absolute or pure relative (interval-based) storage of pitches in memory. This mixed model of intonation is in accordance with a theory of latent AP (and the results of Olthof et al., 2015), in which only the tonic (or another reference pitch, such as the starting or the lowest tone) is stored in an absolute fashion, whereas pitch content is represented in memory and produced relative to this reference pitch [2].

Participants generally kept their tonic pitch between the three different conditions -- only in 5 cases could deviations of more than 1 semitone be observed. One case was an octave shift, another case comprised an upward jump of a fifth (representing an adjustment to a comfortable pitch range), while the other three deviations were all approximately 1.5 semitones. The median change in absolute pitch level between conditions was -0.003 semitones. Hence the participants were very consistent between trials, in agreement with results reported by Bergeson & Trehub (2002) and Halpern (1989).

Finally, the absolute value of the tonic pitches (using rounded MIDI pitch based on A4=440Hz, i.e., C4 = 60) was measured (cf. Figure 1). In both cases (pitched and unpitched click track, see below), the distribution of tonics is clearly not uniform and apparently bimodal, with modes separated by a fifth (Gb-Db in the pitched, and E-B in the unpitched case). Rayleigh tests showed a significant deviation from uniformity in both cases (pitched: *p<.000*, unpitched: *p=.03*). Unfortunately, as realized later, the recordings of one of the two experimenters were made using a (vaguely) pitched click track, which might have primed the participants, even though the click sounds are very brief and the true pitch height is rather hard to perceive. Spectral analysis revealed that the clicks are narrow band noise around a peak of 835 Hz (G#5). Hence, this result needs to be taken with care. However, this might be an interesting finding in and of itself, if these very brief stimuli are actually able to prime singers in terms of their absolute pitch choices. Moreover, the experimenters quite often sung themselves parts of "Happy Birthday" for demonstration purposes while instructing the participants. This might also have biased the choice of absolute pitch height by the participants.
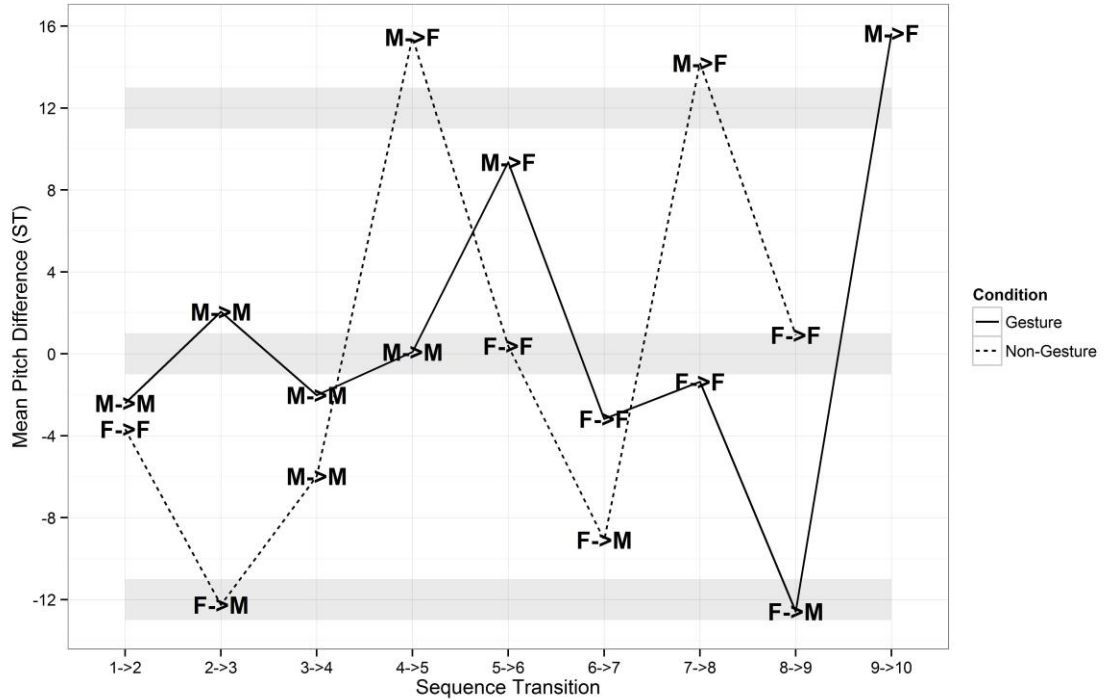
Fig. 2. Differences of mean pitch measured in semitones between consecutive versions in each transmission line. Mean pitch serves here as a proxy for the tonic. Labels indicate the gender of teacher and and students (M=Male, F=Female). The gray areas indicate the [-1, +1] ST intervals around the octaves of the original pitch.

There are two possible interpretations of the results: Either singers can be primed very easily, even with very short click sounds or casual singing samples by an experimenter, or there are actually still undiscovered preferred choices for tonics on which "Happy Birthday" is normally sung [3]. Both possibilities would indicate that (latent) APM does play a role in oral transmission or is a result of vocal and melodic range interactions, as will be seen later. Of course, the possibility that these tonic distributions are purely random cannot be entirely ruled out on the basis that the present dataset is rather small.

ARTIFICIAL ORAL TRANSMISSION

In the second part, I will report on an experiment designed to directly observe the process of oral transmission of a folksong in the lab. Results partly support and partly run against the results of Olthof et al. (2015). Particularly, the results on the trajectory of the tonic during the transmission experiment will be used for the model in the third part of the commentary.

Method

In an experiment carried out during the ISSSM 2011 in Jyväskylä, a group of students set up two artificial oral transmission chains, under supervision of the present author (Asano, Andean, Esser, Gagliardi, Khorsandi, Lenz, Revegno, & Varelmann, 2013). The original hypothesis was that accompanying gestures might facilitate memory recall and learning of simple folk songs [4]. A French folk song ("Cadet Rousselle", 6/8 time in G major) of medium complexity that was previously unknown to the participants was chosen as a stimulus, for which new English lyrics were created as well as a set of accompanying simple gestures. Starting from an initial seed provided by the experimenters, each participant had to learn the song first from the previous participant, then make a solo recording of the song, and finally teach the song to the next participant. No time limit was given. For further analysis, all learning sessions were audio recorded, and, in the gesture condition, video recordings were also made. Next participants in each line were chosen at random, mostly from attendants of the summer school. Each transmission line consisted of

10 participants, however, the last singer in the non-gesture condition had to be excluded, since she was not able to learn the song properly, leaving N=19 valid cases (10 females).

Results

The final recordings of each participant were transcribed manually into standard musical notation by the experimenters, whereas the lyrics were transcribed and analyzed separately. A computational similarity analysis (Müllensiefen & Frieler, 2004) of the melody variants revealed that the meter and rhythm as well as the overall phrase structure were very stable, but that the pitch content was highly volatile except for certain skeleton notes. Contrary to the initial hypothesis, the accompanying gestures did not facilitate but actually hindered learning, as significantly more changes occurred in the gesture transmission line than in the non-gesture chain.

Analysis of the teaching recordings revealed that in all teaching situations instructor and student sung the tune together, normally more than once, and single parts of the song were often repeated many times in unison. However, the final recording of the student's version was done without the teacher being present.

One would expect, from the results of Olthof et al. (2015), that the absolute pitch height should be relatively stable in both transmission lines. However, this was not always the case (cf. Figure 2). The plot shows differences in mean pitch from each transmission step in both chains. Mean pitch serves here as a proxy for the tonic, since the tonic was not always clearly determinable because some singers dropped in overall pitch [5]. As one can see, only three transitions keep roughly the same tonic (all of these pairs are of the same gender), whereas two transitions, both from a female teacher to a male student, resulted in an octave drop. When the opposite student-teacher relationship was assumed (male teacher, female student), the transpositions often over- or undershot the octave by at least 2 ST. All other transitions are in range of a minor or major third or a sixth up and down, only the transition 3→4 in the gesture line is a fourth down (cf. Figure 2). These results are rather striking, particularly in light of the fact that, while all teaching pairs sung the tune together on the same pitch, the students often chose a different tonic for the solo rendition of the song (and probably for subsequent teaching as well).

CONSTRAINED RANDOM WALK WITH SWITCHING STATES

In the third part of my commentary, I will use the results from the other two parts for an attempt to model the transmission of absolute pitch class during oral transmission of folksongs. This can viewed as a more elaborate estimate for the baseline probabilities as used by Olthof et al.

Method

The results of the oral transmission experiment raises the question as to how the observed results for the inter-recording pitch consistency of Olthof et al. (2015) could possibly come about. Even if latent AP is widespread, much of the existing evidence suggests it is far from perfect (Schellenberg & Trehub, 2003; Frieler et al., 2013). Hence, every transmission step will introduce some error in terms of absolute pitch. The simplest model for this is a random walk in pitch space, with some additional constraints introduced by the limited vocal range of a singer. Contrary to Olthof et al (2015), who also calculated a baseline using Monte Carlo simulation, it was assumed that the whole range of a tune is used by a singer to determine the absolute pitch level. In the experiments reported in Mauch et al. (2013), only one case occurred in which a singer started the first trial on a "wrong" starting pitch such that the melody did not fit comfortably into her vocal range, but all other singers, even the unexperienced and "poor" singers, used one single and, presumably, comfortable tonic throughout [6].

Let's consider the following basic model for the lowest pitch of a folk song of the form $p_i \rightarrow p_i + \delta(p_i, g_i) = p_{i+1}$, where the error is either normally distributed $\delta \sim N(0, \sigma)$ for one transmission step with a mean of 0 and standard deviation σ (measured in semitones) or uniformly distributed in the available pitch space. This error term is thought to comprise errors of recall of absolute pitch height. The smaller the standard deviation, the more exact the recall of AP. The uniform error, on the other hand, models fully unreliable APM (limit $\sigma \rightarrow \infty$).

Next, introduce a gender state $g_i \rightarrow g_{i+1}$ with fixed probabilities $p(g=F) = p_F$, $p(g=M) = 1\text{-}p_F$. Furthermore, assume a fixed vocal range $V(g) = [p_{min}(g), p_{max}(g)]$ for all singers of the same gender in the transmission line. Olthof et al. (2015) report a typical range for male singers of roughly one octave and a tritone (G#2-D4), and for female singers of one octave and a fifth (F3-C5, or 53-72 in MIDI pitch). For symmetry reasons, the male range was slightly extended to G2-D4 (43-62 in MIDI pitch). Next, assume that the melodic range of a song stays constant during transmission, and denote the ambitus in semitones with *A*. The effective vocal range for the lowest tones is then restricted to $V_{eff}(g) = [p_{min}(g), p_{max}(g)\text{-}A] = [p_{min}(g), p_{max}'(g)]$, if the singer wants to fit the melody into her vocal range. This means, that melodies can have at most an ambitus of 19 semitones in this model [7]. Furthermore, introduce a strategy for the case that the new randomly generated pitch would fall out of the vocal range, which can happen particularly in the case of transmission between singers of different genders, but also just by chance, if the current pitch is close to the lower boundary of the vocal range. In the case of Gaussian error, one simple remedy consists in forcing the error for each step into the interval [$p_{min}\text{-}p_i$, $p'_{max}\text{-}p_i$] by re-generating error values. In the case of uniform noise (i.e., absence of APM) a value from this interval is chosen at random. For gender switching (no pun intended), the following strategy was adopted, inspired by the data from the oral transmission experiment: Random jumps of one octave, one fifth or one fourth up or down were generated successively till the pitch level fell into the current effective vocal range. Afterwards the normal error term was added. A probability distribution for this was arbitrarily fixed to be $p(J=octave) = \frac{1}{2}$, $p(J=fifth) = 1/3$, $p(J=fourth) = 1/6$. The jump interval is a function of the current pitch and genders $J = J(p_i, g_i, g_{i+1},)$. For a same-same-gender transition the jump interval is set to 0, in the other two cases, intervals will be chosen at random and added up until the sum of current pitch and jump interval falls into the vocal range of the next singer, i.e., until $p_i + \Sigma J_k \in V_{eff}(g_{i+1})$. The full model is then

$$p_i \rightarrow p_i + \delta_M(p_i, g_{i+1}) + J(p_i, g_i, g_{i+1},),$$

$$g_i \rightarrow g_{i+1},$$

where the error term δ_M can be either the clamped Gaussian or the uniform distribution. For the case of Gaussian error, three different values of sigma $\sigma \in \{.25, 1.0, 1.5\}$ were used. Furthermore, the melodic range was varied between 5 and 16 semitones. Gender probability was fixed to the value $p_F=0.8$ as used by Olthof et al (2015). For each combination of parameter settings, $R=10$ tune families of $N=60$ songs each were simulated using transmission lines of lengths $L \in \{3, 30, 300\}$. The starting pitch for each tune family was set in the mid-point of the effective vocal range plus a random offset, sampled from a Gaussian distribution with a mean of zero and the same standard deviation σ as the employed Gaussian error. In the case of uniform error, the starting pitch was chosen uniformly in the effective vocal range. This assumes that all melodies in a tune family started from a common original source. Finally, the tonic pitch was determined by adding a randomly chosen constant to the lowest pitch. The constant was sampled from the distribution of intervals between lowest and tonic pitches in the Luxembourg/Lorraine collection, which primarily have the either the dominant (54% of songs) or the tonic (22% of songs) as the lowest pitch. For each of the resulting distributions of tonic pitches (mapped to pitch classes), a Rayleigh test of uniformity was conducted with a fixed significance level of $\alpha=.05$ [8].

Results

Results of the simulations can be found in Figure 3. The most important observation is that beyond a critical melodic range of 10 semitones, all Rayleigh test become significant, independent of all other parameters, notably from type of error, which represents the influence of APM. If the proposed model indeed captures some true aspects of oral transmission processes this means that, in contrast to the conclusion of Olthof et al. (2015), the limited vocal range of the singers might be the main reason that distributions of tonics deviate significantly from uniformity. This is particularly important with respect to the fact that the median pitch range of the Luxembourg/Lorraine song collection is exactly one octave, while about 85% of the songs have an ambitus of 10 or more semitones, i.e., most of the tunes fall beyond this critical threshold [9]. Thus, if limited vocal range is the main reason for the observed non-uniform distributions, how can one differentiate and extract possible true APM effects? One possibility comes from the results for melodies with a very small melodic range (cf. Figure 3), because these produce primarily

uniform pitch class distributions in the case of uniform noise (no APM), but not in the case of Gaussian noise (APM).

CONCLUSION

In this commentary, I proposed a more elaborate model for the transmission of the tonic, which suggests that the results of Olthof et al. (2015) might partly be explainable by an interaction of melodic and vocal range. However, this model still implies that some form of latent APM plays a role in this process, but with the modification that latent APM is mediated (or even caused) by limited vocal ranges of singers. The model is, however, still rather simple and tentative, and relies on several plausible but untested assumptions. For instance, both the melodic range within a tune family and the vocal range of the singers were assumed to be constant. An improved explanation would use more flexible models, which on one hand should be based on data taken from the songs themselves, and on the other hand should incorporate experimental data on vocal ranges. Still, it seems clear that the interplay of melodic and vocal ranges might play an important role in influencing the tonic distributions to deviate from uniformity, which hence cannot reliably interpreted as the influence of latent APM alone. But, as the simulation shows, investigating tune families with very small melodic ranges could shed light on this issue.

Furthermore, relying on the Rayleigh test exclusively might not be sufficient. Considering the case of LOOS (Olthof et al., 2015: Figure 1), the distribution of tonic pitches is heavily concentrated around two pitch classes, which is a highly improbable configuration, if the melodic range in this tune family is not unusually large (unfortunately, Olthof et al. (2015) do not report statistical properties, such as melodic range, of the tunes), and might indeed indicate a true influence of APM. Perhaps special measures of the concentration of circular histograms are needed to discriminate these cases from merely non-uniform distributions. However, these tests are yet to be developed. Using normalized circular entropy in combination with Rayleigh tests might be one fruitful step in this direction.

The results of our simulations show at least that the interaction of melodic range and vocal range cannot be fully ruled out. Hence, the alternative hypothesis of pitch class uniformity might not be the most appropriate alternative hypothesis, if the interaction of melodic and vocal range constrains the possible pitch classes *a priori*. Likewise, bi-modality could still be a sign of latent APM, as seen from the two experiments reported here, since a jump of a fourth or a fifth might be a common adaption strategy to accommodate for a comfortable pitch height.

On the other hand, even if the results can be fully or partly attributed to this interaction, this still requires that singers are able to remember the most comfortable pitch height for a well-known song. Or, speculating further, tunes falling into one's comfortable range might even be remembered better (e.g., by external or internal singing along).

Finally, some melodic modifications observed in oral transmission might also be the result of adapting a tune to one's vocal range while trying to retain the absolute pitch height, e.g., by modifying single very high or very low peaks which fall outside one's comfort zone.
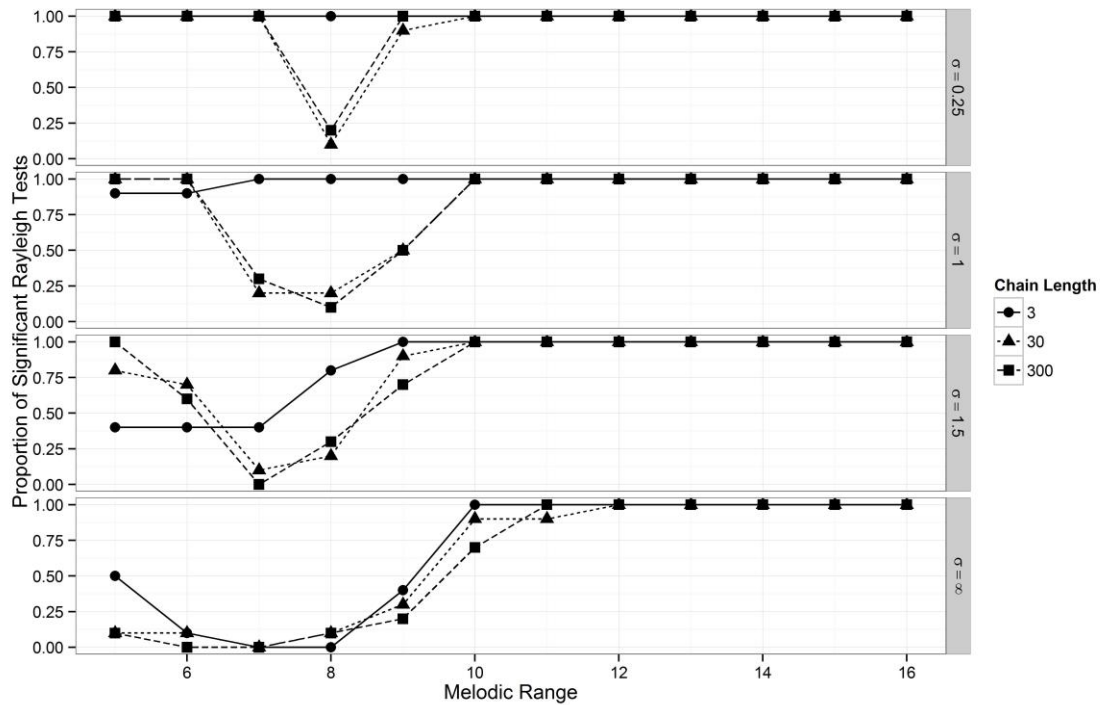
Fig. 3. Mean value of proportions of significant Rayleigh tests ($\alpha=.05$) for simulated circular distributions of tonic pitches as a function of melodic range (ST). Top to bottom panels: Clamped Gaussian error with $\sigma \in \{0.25, 1.0, 1.5, \infty\}$, where $\sigma=\infty$ represents uniform error. Simulations were done for 10 tune families with 60 songs each, using three different transmission chain lengths $L \in \{3, 30, 300\}$ and constant gender probability $p_F =.8$.

Outlook

There are still many open questions regarding the specifics of oral transmission in folk song practice. This applies to both the specifics of singing and intonation strategies, as well as to the necessity of improvements in the modeling of latent APM. A tentative list follows:

1. Investigate the strategies that singers of all skill levels employ in order to fit a given melodic range into their vocal ranges. Are there preferred zones within one's vocal range? Is the assumption of an average vocal range for each gender justified, or would the system of classical voice types (soprano, alto, tenor, bass etc.) be more adequate?
2. Further scrutinize the stability of APM for well-known songs in a longitudinal study.
3. Conduct further oral transmission experiments in the lab to gain deeper insight into the details of the processes, using longer chain lengths and one-to-many teaching/learning.
4. Relate the observed absolute pitches to musical properties of the tunes such as ambitus and interval distribution.

To sum up, the exciting research presented by Olthof et al. (2015) opens up interesting questions in many directions, which are definitely worth further pursuing.

END MATTER

Acknowledgements

NOTES

 [1] Now freely available in the Tony software (Mauch et al. 2015) at https://code.soundsoftware.ac.uk/projects/tony.

[2] One follow-up question is then, if the mixed model for tune memory holds also for music for which a single unequivocal tonic cannot be assigned, e.g., heavily chromatic, modulating or 12-tone melodies.

[3] An interesting experiment to address the latter hypothesis would be to record renditions of "Happy Birthday" in the wild and analyze the chosen tonics. Connected to this is another interesting question: How does a random birthday choir choose and decide on a common tonic (if they do)? Likewise, Kopiez & Brink 1998) observed that fan chants often show fixed absolute pitch height.

[4] This hypothesis was inspired by alleged Finnish folk practices of singing and learning the Kalevala, the famous Finnish national epic.

[5] This can be justified by the facts that a) the mean pitch of our stimulus folk song lies between the tonic and the second scale degree, and b) the melodic variations left the overall pitch range mostly intact.

[6] In contrast to our assumption, Olthof et al. (2015) consider only the first five notes of a melody and the first tonic to simulate their baseline. But the range of the first five pitches is often considerably smaller than the whole range of a melody (the median pitch range for the first bar in the Luxembourgian and Lorraine songs of the Essen collection is 4 ST). Actually, both assumptions are justifiable, so more empirical tests would be needed to decide these matters. However, it seem more likely to us that a singer performing a well-known song tries to fit the whole song instead of only the beginning in the most comfortable zone of her vocal range.

[7] For the modeling and the following discussion I will use the Luxembourg and Lorraine sub-corpus of the Essen collection with N=1092 songs (Sagrillo, 1999; Schaffrath, 1995) as a benchmark. I assume that the data extracted from these collections are quite similar to the Dutch folk songs use by Olthof et al. (2015). All features were extracted using the MeloSpyGUI (Frieler, Abeßer, Zaddach & Pfleiderer, 2013a), available for free download at http://jazzomat.hfm-weimar.de. The range of pitch ranges in this set of melodies is 4-20 ST; with a median of 12 ST, while 50% of the tunes have a range between 12 and 14 ST.

[8] I did not apply Hartigan's dip test, since I have doubts that this test is applicable to circular data, because the derivation of this test relies heavily on distribution functions on the real line, which are not defined for circular variables. Alternatively, the maximum half-split ratio was calculated, which is defined by dividing the circular histogram into all possible two halves and calculating the ratio of frequencies in both halves. The maximum over all possible axes is the maximum half-split ratio (MHSR). If the distribution is uniform this ratio should be close to .5, the higher the concentration of values in one half, the higher is the MHSR. As it turns out, Rayleigh test p-values are highly correlated with the MHSR (Spearman's $\rho=.89$, $p<.000$), so I do not show the distributions.

[9] Clearly, as indicated by the curious dip at about 8 ST in Figure 3, particular in the case of very good APM ($\sigma=.25$), this threshold of non-uniformity is a function of melodic range and vocal range, depending critically on these, thus, on model details, which needs further scrutiny. Hence, the exact numerical values of 8 and 10 ST should be taken with care.

REFERENCES

Asano, R., Andean, J., Esser, H. H., Gagliardi, M., Khorsandi, A., Lenz, C. Revegno, R. & Varelmann, T. (2013). Report - The International Summer School in Systematic Musicology. In: Wewers, J. & Seifert, U. (Eds.): *Under Construction: Trans- and Interdisciplinary Routes in Music Research. Proceedings of SysMus 11*, Cologne 2011

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

Frieler, K., Abeßer, J., Zaddach, W.-G., & Pfleiderer, M. (2013a). Introducing the Jazzomat Project and the Melo(S)py Library. In: Kranenburg, P. van; Anagnostopoulou, C.; Volk, A. (Ed.) *Proceedings of the Third International Workshop on Folk Music Analysis*, Meertens Institute and Utrecht University Department of Information and Computing Sciences, pp. 76–78.

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

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

Hartigan, J. A. & Hartigan, P. M. (1985). The Dip Test of Unimodality The Annals of Statistics, 13(1), 70–84.

Kopiez, R. & Brink, G. (1998). Fußball-Fangesänge. Eine Fanomenologie. Würzburg: Königshausen & Neumann.

Mauch, M., Frieler, K. & Dixon, S. (2014). Intonation in Unaccompanied Singing: Accuracy, Drift and a Model of Reference Pitch Memory. *Journal of the Acoustical Society of America*, 136(1).

Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Dai, J., Bello, J. & Dixon, S. (2015). Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency, in: *Proceedings of the First International Conference on Technologies for Music Notation and Representation*.

Müllensiefen, D. & Frieler, K. (2004). Cognitive Adequacy in the Measurement of Melodic Similarity: Algorithmic vs. Human Judgments. *Computing in Musicology*, 13, 147–176.

Olthof, M., Janssen, B. & Honing, H. (2015). The Role Of Absolute Pitch Memory In The Oral Transmission of Folksongs. *Empirical Musicology Review*, this issue.

Sagrillo, D. (1999). *Melodiegestalten im luxemburgischen Volkslied: Zur Anwendung computergestutzter Verfahren bei der Klassifikation von Volksliedabschnitten*. Bonn: Holos.

Schaffrath, H. (1995). The Essen Folksong Collection in Kern Format. In Huron, D. (Ed.), *Computer Database*, Menlo Park, CA.

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