

7 A Hidden Markov Model of Melody Production in Greek Church Chant

Panayotis Mavromatis

Department of Music and Performing Arts

New York University

35 West Street, Room 682-B2

New York, NY 10012-1172

panos.mavromatis@nyu.edu

Abstract

We present a probabilistic model of melody in modern Greek church chant. The tradition relies on memorization and improvisation skills which are taught without explicit appeal to rules. The researcher is faced with the challenge of inferring the rules of the idiom from a sample corpus. The structure of the rules will point to the mental representation of melody that underlies learning, recall, and improvisation. Our analysis is performed in two stages. In the first, a Hidden Markov Model (HMM) is trained on a corpus of chants, using a variant of the state-merging algorithm of Stolcke and Omohundro. In the second stage, the optimum HMM is analyzed. Its states can be viewed as probabilistic rules that determine the course of a melody, given its preceding melodic and textual context. Our findings show that, given the pattern of textual word-stresses and syntactic groupings, the shaping of the melody within a given mode (Echos) is completely determined by a small number of phrase parameters reflecting melodic choices at key decision points. We discuss our model in relation to previous work and in the cognition of melody.

Music Analysis East and West (Computing in Musicology 14).

Published by CCARH and The MIT Press, 2005-06.

7.1 Modeling Orally Transmitted Skill in Greek Chant

This study develops a mathematical model of melody in Greek church chant, the music of the Greek Orthodox service as practiced nowadays in the Greek mainland and in communities around the world. This music descends from medieval Byzantine chant; it is traditionally purely vocal, monophonic, and commonly employs a moving drone.

Our formal modeling approach is motivated in part by an intriguing aspect of the Greek chant tradition: the ability of chanters to learn and/or transmit the melodic style orally, without recourse to explicit rules, or even to musical notation. This skill is clearly necessary for those who cannot read notation, or who do not have access to chant books. But even chanters trained in musical notation are able to improvise when exact rendering of a chant is not called for. It appears that, even at different levels of skill, chanters develop an internalized knowledge of the melodic idiom, unconsciously extracting rules and patterns from the concrete examples of melodies they encounter. The phenomenon is similar to natural language acquisition. A formal model of Greek chant melody can therefore be viewed as a generative grammar that reflects a chanter's internalized knowledge of the style, or, to borrow a term from linguistics, his or her "competence" (Chomsky 1965).

Our task will be to construct a grammar of Greek chant based on melodies recorded in printed chant books. The main challenge is that the rules of chant melody are too complex to write down directly. We will need to use an algorithmic method, within a framework computer scientists often refer to as *grammatical inference*. This motivates us to consider Hidden Markov Models, a technique widely used to analyze sequential data. Roughly speaking, a Hidden Markov Model (HMM) is a probabilistic version of a Finite-State Machine; the latter represents a grammar graphically as a collection of paths in a melodic space, capturing all grammatical realizations in a compact and visually intuitive way. HMMs add probabilities to each path. Thus they allow us to represent variable outcomes in the improvisation process and to rate each melody according to its typicality. In addition, there are algorithms for building an HMM from a corpus of examples. The shape of the resulting model can suggest probabilistic rules that determine the course of a melody given its preceding melodic and textual context. It is then straightforward to compute the set of all possible melodies corresponding to a given text.

If we collect together all melodies assigned to a given chant text by one of our HMMs, an interesting picture emerges: given the word-stress pattern, the melody can be completely determined by a relatively small number of note choices made at key decision points. Each melodic phrase typically contains two to three such points, which divide it into formulaic chunks. Each chunk is about six to nine notes long, with a melodic contour that highlights word-stress. The decomposition of the melody into "chunks" is consistent with psychological theories of sequence

representation (Miller 1956; Cowan 2001), which suggest its purpose may be to reduce processing load in real-time tasks such as recall and improvisation. Our work therefore not only makes explicit the formulaic structure of Greek chant, but it also indicates how this structure may originate in properties of orally transmitted skill.

In Section 7.2, we illustrate the melodic style under investigation through examples, identifying factors that shape its melody. We explain the use of Finite-State Machines as an efficient way to capture these factors. The problem of building the grammar from a corpus is considered in Section 7.3, where HMMs are introduced and the grammar-learning algorithm is explained. In 7.4 we describe the structure of the resulting solutions in terms of decision points and chunks. In Section 7.5, we discuss the relation of our research to previous work in chant studies and in the cognition of melody, pointing to directions for future research. (For audio examples of Greek chant, see www.goarch.org/access/byzantinemusic/.)

7.2 Modeling Melody with Finite-State Machines

7.2.1 The Melodic Style of Greek Chant

Each chant in the Greek service receives a fixed modal designation, known as its *Echos* (pl. *Echoi*). The eight Greek *Echoi* fall into the authentic/plagal scheme also employed in Western chant. It should be noted that the Greek and Western modal systems of chant are structured and conceptualized in different ways, and few of the eight Greek *Echoi* bear any tonal resemblance to their Western counterparts. The modern theory of the *Echoi* originates in Chrysanthos (1832). For more recent accounts, see Desby (1974), Giannelos (1996), Karas (1985), and Mavroeidis (1999).

The *Echos* determines the chant's melodic character by specifying a scale, tonal centers, and characteristic melodic patterns. In the oral mode it is important to have internalized the *Echos's* elements and to be able to apply them when realizing a chant without notated melody. Of course, this knowledge need not be verbalized. In fact, the theory of the *Echoi* is only explicit about scales and *principal tones*; the latter include the *Echos's* final and a set of cadential pitches of varying strengths. The theory also tells us that melodic phrases, as defined by cadences, end at textual syntactic boundaries, which are typically marked by punctuation.

We will focus this discussion on a single tonal/formulaic environment, namely *Echos* 1 in its syllabic form. This will be adequate for the present purpose of illustrating our computational model. Similar analysis pertains to the remaining *Echoi*. In the Greek system, a chant belongs to one of three broad stylistic categories which are differentiated by relationships between text and music. These categories are called syllabic (*heirmologikon*; one note to a syllable), neumatic (*sticherarikon*; multiple notes to a syllable), and melismatic (*papadikon*; many notes to a syllable). Membership affects not only the *Echos's* formulaic content but also, in many cases, its tonal

centers. There are about twenty different formulaic environments that can be analyzed in the manner described in this article. An analysis of the Greek modal system in its entirety is beyond the present scope and will be pursued in a future publication.

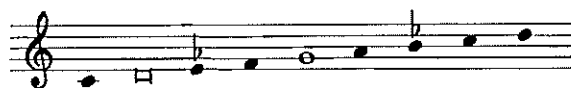


Figure 7.1. The scale of *Echos 1*, with principal tones shown in white noteheads; the breve [rectangular notehead] indicates the *Echos's* final.

Figure 7.1 summarizes the theoretical characterization for *Echos 1*, while Figure 7.2 shows typical chants in that *Echos*. Note that Greek chant notation is neumatic and encodes *melodic intervals* rather than absolute pitches. Here we will use a standardized system of transcription, but it should be understood that this is only an approximation, as Greek scale tunings differ considerably from those of the West.

The static characterization of Figure 7.1 does not fully address the dynamic character of the *Echos* as seen in Figure 7.2. First, the *Echos* involves characteristic melodic formulas, which are especially prominent at phrase endings, sometimes also at phrase beginnings. Second, word stress appears to affect the choice of melodic intervals: a stressed syllable of text is almost always preceded by an ascending interval, or followed by a descending one, or both. There are very few exceptions to this pattern, and these occur mostly in the context of formulaic phrase endings or beginnings. One of our main goals will be to capture such pitch-stress relations, showing how they interact with the tonal framework of the *Echos* to determine its formulaic, as well as its non-formulaic, material.

(a) $\overset{\text{CF_G01}}{\text{—}} \overset{\text{CF_D01}}{\text{—}}$
 Ἀ γι ὦ Πνευ μα τι τι μη και δο ξα ως περ Πα τρι πρε πει α μα και Υι ὦ
 $\overset{\text{CF_G02}}{\text{—}} \overset{\text{CF_D01}}{\text{—}}$
 δι α του το α σω μεν τη Τρι α δι μο νο κρα το ρι α

(b) $\overset{\text{OF_01}}{\text{—}} \overset{\text{CF_G01}}{\text{—}}$
 Ἀ γι ὦ Πνευ μα τι πα σα η Κτι σις και νουρ γει ται
 $\overset{\text{CF_D01}}{\text{—}} \overset{\text{OF_01}}{\text{—}} \overset{\text{CF_D01}}{\text{—}}$
 πα λιν δρο μου σα εις το πρω τον ι σο σθε νες γαρ ε στι Πα τρι και Λο γω

(c) $\overset{\text{OF_01}}{\text{—}} \overset{\text{CF_G01}}{\text{—}} \overset{\text{CF_G02}}{\text{—}}$
 Ἀ γι ὦ Πνευ μα τι τι μην προ σκυνη σιν δο ξα και κρα τος ως Πα τρι τε α ξι ον
 $\overset{\text{CF_D01}}{\text{—}} \overset{\text{CF_G01}}{\text{—}}$
 και τω Υι ὦ δει προ σφε ρειν μο νας γαρ ε στιν η Τρι ας
 $\overset{\text{CF_D01}}{\text{—}}$
 τη φυ σει αλλ ου προ σω ποις

Figure 7.2. Typical syllabic chants in Echos 1. Dots above the Greek text represent levels of word stress, with double and single dots representing stressed and unstressed syllables respectively. The brackets above the staff identify melodic formulas; their labels are explained in Section 7.4.

1
 Του λι θου σφρα γι σθεν τος υ πο των Ι ου δαι ων

2
 Του Γα βρι ηλ φθεγ ξα με νου σοι Παρ θε νε το χαι ρε

3
 Εν τω θλι βε σθαι με ει σα κου σον μου των ο δου νων

4
 Α γι ω Πνευ μα τι πα σα η φυ σις και νουρ γει ται

5
 δε ξι α σου χει ρι λα βων συ Λο γε

6
 Τους ε ρη μι κοις α παύ στος ο θει ος πο θος εγ γι νε ται

7
 Α γι ω Πνευ μα τι τι μη και δο ξα ωσ περ Πα τρι

8
 εις τα ο ρη των σων υ ψω σας με νο μων

Figure 7.3. A phrase family in *Echos 1*, illustrating formulaic variation. Brackets above the staff mark the family's opening and closing formulas.

To get a closer look at formulaic variation, let us examine the group of phrases shown in Figure 7.3. These phrases, coming from different *Echos-1* chants, appear to be variations of the same basic scheme, as highlighted by their alignment. The following features seem to characterize the phrase family as a whole:

- The body of the phrase, enclosed by the first and third vertical dashed lines, revolves around G, one of the principal tones of the *Echos*. The motions away from G and back are presumably caused by different levels of word stress.
- An optional opening formula, marked by the first square bracket above the staff, outlines the pitch sequence D E F G, and serves as an initial ascent to G, starting from the final D of the *Echos*.
- A cadential formula, marked by the second square bracket above the staff, outlines the pitch sequence G F E F G, leading the phrase to a cadence on G.

Both the opening and cadential formulas alter their basic pitch sequence through deletions or duplications, seemingly caused by the different stress patterns of the underlying text.

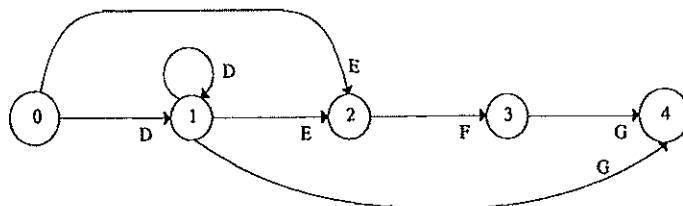


Figure 7.4. A Finite-State Machine for the opening formula of Figure 7.3. The outputs are letter names for the generated pitches.

7.2.2 The Finite-State Machine Representation

A Finite-State Machine (FSM) provides a concise representation of insertions and deletions such as the above, and can be nicely visualized as a graph. An FSM is formally defined by states and transitions, graphically represented by circles and arrows respectively. Figure 7.4 shows an FSM with five states, labeled 0 through 4. Each FSM transition produces an output symbol—in our example, one of the letters D, E, F, G, which represent pitches. An FSM is also characterized by an initial state (here State 0) and one or more final states (here State 4). To generate a string of symbols, we simply choose a path connecting the initial state to a final state, following the direction of the arrows; the generated string is the ordered sequence of output symbols encountered in the path. In our example, going through the states 0, 2, 3, 4 produces the sequence E F G, corresponding to the opening of Figure 7.3, Phrase 5.

The reader can verify that all instances of the opening formula in Figure 7.3 (Phrases 1–7) can be generated from the FSM in Figure 7.4. The advantage of using an FSM may not be immediately obvious. After all, the melodic pattern just examined can be described in words fairly easily, if somewhat awkwardly, perhaps along the following lines: “one or more Ds, optionally followed by E, F, followed by a G; when the E, F is present, the initial D may be omitted.” But the simplicity and elegance of the FSM representation will become clear once we consider more complicated melodic patterns, such as the cadential formula of Figure 7.3. The reader may verify that the FSM of Figure 7.5 generates all instances of that formula. Any attempt at an equivalent verbal description of the pattern will be too awkward to use.

How we choose to represent a melody in an FSM is often a matter of convenience, as long as no relevant information is lost. In fact, we will adopt a representation that facilitates interpretation of an FSM in terms of real-time melodic process: a transition will represent an event triggering some change. A state will likewise correspond to a period of time in which there is no change at the level of detail we are considering.

Since output symbols are associated with transitions, hence with points in time, it will be more natural to choose outputs to represent *intervals* rather than *itches*. Pitches characterize time spans, and are therefore more naturally associated with states. This interval-based representation of chant melody has proven most convenient and can be always converted to a pitch-based one if needed, given a scale and final as determined by the *Echos*. (Since Greek chant notation is interval-based, one could say that the note-based representation shows the influence of Western practice.) Figure 7.6 shows a straightforward modification of the FSM of Figure 7.5 that implements the interval-based representation.

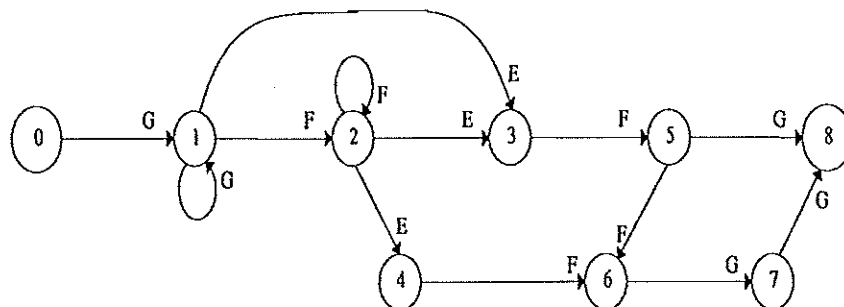


Figure 7.5. A Finite-State Machine for the cadential formula of Figure 7.3. The pitch representation is the same as that of Figure 7.4.

The FSMs we have considered so far model pitch patterns alone. A simple extension of an FSM, known as a *Finite-State Transducer* (FST), allows us to model the patterns of text-pitch relations that are so crucial in shaping the melody. Instead of just output symbols, the FST transitions produce input/output *pairs*. When traversing an FST, we must follow a path whose input symbols match those of the input sequence—in our case the stress pattern of the words; the output symbols collected along the path will then represent the generated melody. Figure 7.7 shows a straightforward incorporation of word-stress inputs into the model of Figure 7.6.

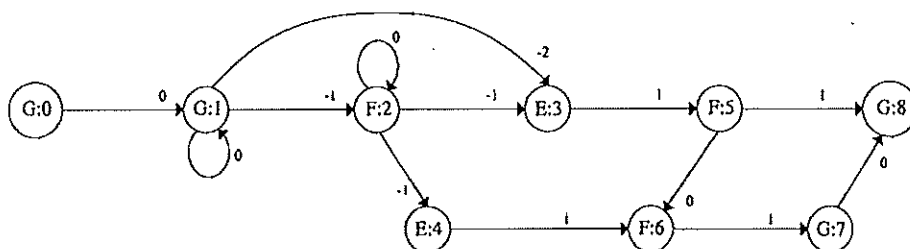


Figure 7.6. A Finite-State Machine for the cadential formula of Figure 7.3 that uses an interval-based representation. Integers stand for intervals measured in diatonic steps, with 0 representing unison and negative integers representing descending intervals. Each state is labeled by its resulting pitch; the choice of pitch for the initial state is arbitrary.

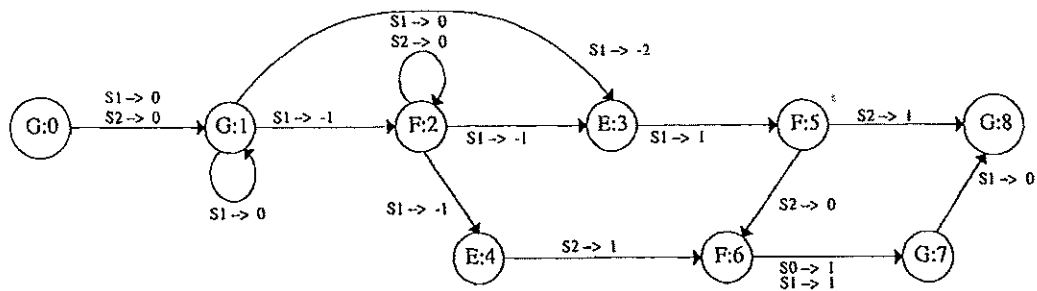


Figure 7.7. A Finite-State Transducer for the cadential formula of Figure 7.3. A pair of input/output values is represented here as input \rightarrow output. The input is word stress, which takes the values S2 (stressed syllable), S1 (unstressed syllable), and S0 (no syllable change). The output is the melodic interval, as in Figure 7.6.

7.3 Hidden Markov Models and State-Merging

7.3.1 Hidden Markov Models

The FSMs presented in the previous section were constructed by simply inspecting the sample corpus of Figure 7.3. However, it is impractical to construct a complete model in this way, since a realistic corpus could involve hundreds of examples. In this section we will show how to obtain the model algorithmically. The algorithm involves a stochastic generalization of FSMs known as Hidden Markov Models (Rabiner 1989; Manning & Schütze 1999: 317–340).

A Hidden Markov Model (HMM) is defined as an FSM with probabilities attached to its transitions and output symbols. Transition probabilities emanating from a given state add up to 1, and so do output probabilities on a given transition. The generation of an output string through a specific HMM path has probability equal to the product of all transition and output probabilities encountered in the process. The “hidden” in HMM comes from the fact that when we observe an output string, we generally do not know the path that generated it. More than one path can produce the same result, and the output string’s overall probability is obtained by summing up the probabilities of all its possible derivations. In an HMM, the states are not postulated explicitly; rather, they are inferred from a corpus through a training algorithm. As such, the states’ meaning may not be obvious, and will have to be retrieved by appropriate analysis. In fact, in many applications the meaning of the HMM-states need not be uncovered at all. The model can be used as a “black box,” i.e., a device that produces the correct result, without us understanding its internal structure.

The HMM training algorithm used in this study originates in Stolcke & Omohundro (1993, 1994). The advantages of the Stolcke-Omohundro (SO) algorithm include its great flexibility in determining the model’s connectivity, and its ability to incorporate constraints on the model’s geometry that may arise from the specific domain under investigation. As we will see, these properties give us some control over the training

process and make it easier to obtain an interpretable model. The SO algorithm starts with a maximal HMM that over-fits the data and proceeds to systematically reduce the model size by merging together states until an optimal model is reached. It will be best to illustrate this process through an example in order to show how the set of realizations of the cadential formula in Figure 7.3 could have led us to the model of Figure 7.7.

Let us first note that the phrase family of Figure 7.3 includes eight phrases but only exhibits *five* different realizations of the cadential formula (Phrases 1, 4, 5, and 8 all share the same realization). We can construct an HMM that generates exactly these five instances as shown in Figure 7.8. We begin with one initial and one final state. For each distinct instance of the cadential formula, we then connect the initial to the final state by a path that generates that instance but shares no other states with the rest of the graph. Each path is assigned a probability according to the multiplicity of the instance(s) it generates. In Figure 7.8, for example, the fourth path from the top is encountered four times more often than the rest. The five paths therefore receive probabilities 0.125, 0.125, 0.125, 0.5, and 0.125 respectively. This is achieved by giving the corresponding probability to each path's first transition; everything else is assigned probability 1.

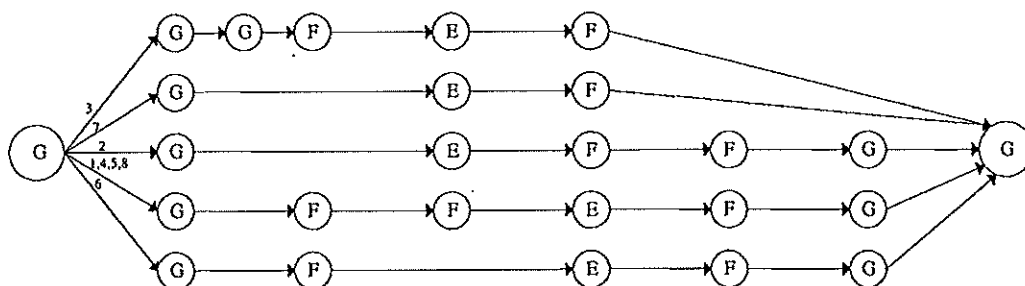


Figure 7.8. The SO algorithm's starting model based on instances of the cadential formula of Figure 7.3. Output symbols are suppressed for simplicity. The initial transition of each path is labeled by the corresponding phrase number(s) from Figure 7.3.

The main problem with the above FSM is that it does not extract any general rules from the concrete observations, since it fails to abstract patterns that may be shared by the instances. The SO algorithm introduces generalization by merging together states that correspond to such shared patterns. For instance, the lower three paths in the model all end in F, G, G in the same context of word-stress and pitch (see Figure 7.3). The three F-states could therefore be considered functionally equivalent, and similarly for the three G-states not shared by the paths. Merging together equivalent states involves combining the sets of transitions in and out of these states, along with their attached outputs. The associated transition and output probabilities are then recalculated as weighted sums of the corresponding probabilities before merging. This calculation is local in that it only affects the merged states and their direct

neighbors. Details can be found in Stolcke & Omohundro (1994). The state-merges suggested by the shared pattern F, G, G lead to the HMM shown in Figure 7.9.

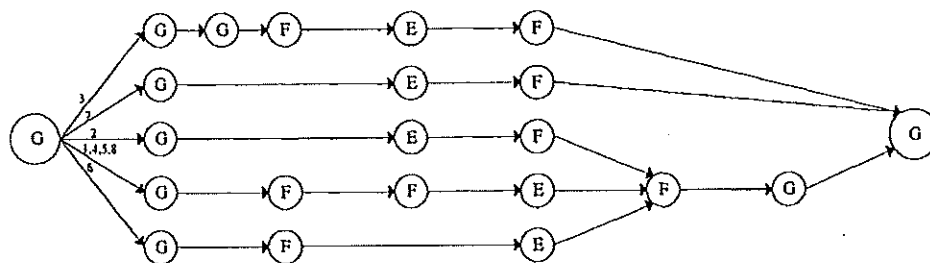


Figure 7.9. The model obtained from Figure 7.8 after a few state-merges identified by a shared pattern.

We continue this process until we capture all relevant shared patterns. In fact, the state-merges shown in Figure 7.10 indeed lead to the model of Figure 7.7. Putting aside for the moment the question of how these merges were chosen, let us observe that, in light of their origin in the merging process, the states in Figure 7.7 have a natural musical interpretation: each state represents a melodic function that is characterized by its pitch as well as its melodic context. For example, state F:2 represents an optional note F, subordinate to its more stable surrounding G and E, like the passing tone of Western theory. By contrast, F:5 and F:6 represent notes necessary to the formula—in the sense that *either one or the other must be present*—and act as a preparation to the cadential G:8. On a finer level of detail, F:5 and F:6 represent different functions, since the former is unaccented, whereas the latter is accented and requires the “anticipation” note G:7. This discussion strongly suggests that an *interpretable* HMM can indeed represent the dynamic aspect of an *Echos*’s pitch structure, moving beyond the static representation of Figure 7.1, as we had hoped.

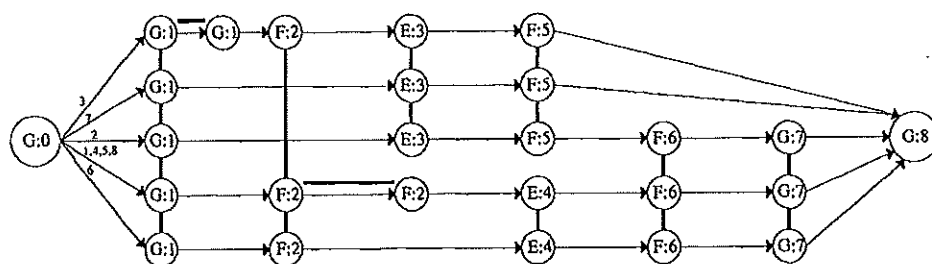


Figure 7.10. The state-merges leading to Figure 7.7. Thick lines connect the states to be merged.

Consider now what happens if we continue merging states beyond Figure 7.10. We may end up with the FSM of Figure 7.11, obtained by merging E:3 with E:4, and F:2 with F:5 and F:6. At this point, we may rightly suspect that the process has gone too far: the model now generates practically any sequence of Es and Fs between the

initial and final Gs, including many sequences that are never observed in practice. In other words, the model *over-generalizes* and thus loses predictive power. Moreover, in merging together states F:2, F:5, and F:6, for example, we lose all the nuance contained in the states' functional differentiation as discussed earlier. The model is no longer straightforward to interpret. It is therefore clearly important to know when to stop merging.

To address this problem, the SO algorithm introduces a "cost" function that numerically evaluates each candidate model. We describe this function qualitatively and show how it can be used to terminate the algorithm. The interested reader can consult Stolcke & Omohundro (1994: 7-27) for more details.

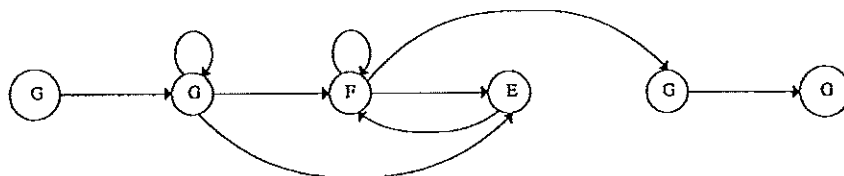


Figure 7.11. Merging states beyond the optimal model of Figure 7.7. This model over-generalizes.

The models of Figures 7.7, 7.8, and 7.11, ordered by the merging process, can be evaluated according to at least two criteria. The first criterion is *goodness-of-fit*, measured by the overall likelihood of occurrence a model assigns to the observations. As states are merged, the observations become a smaller part of all generated sequences, hence goodness-of-fit decreases. In this respect, Figure 7.8 shows perfect fit, since it generates just the observations. As we saw earlier, we must give up some of that fit for the sake of generalization. A second criterion that contrasts the three models is *simplicity*, which depends on the model's size and connectivity. State-merging pushes towards greater simplicity, with Figure 7.11 representing the simplest model of the three. Again, we must give up some of that simplicity to avoid over-generalization.

Since goodness-of-fit and simplicity represent opposite forces, it may be natural to combine them, so that an optimal model embodies the best balance between the two. This is achieved through the SO cost function, which quantifies lack of goodness-of-fit and of simplicity as *model error* and *model complexity* respectively:

$$SO \text{ cost function} = \text{model error} + \text{model complexity}$$

The problem then becomes choosing the merges that will minimize this function; we can stop merging when the minimum cost is achieved. Indeed, in the early part of the SO algorithm, model complexity dominates. Model cost is high and decreases as we merge states. Once the minimum is reached, model error begins to take over, causing the cost to rise again. In our case, the lowest-cost model turns out to be that of Figure 7.7.

7.3.2 Choosing Which States to Merge

A complete model of an *Echos* typically involves hundreds of states. In such a graph, searching for the best sequence of state-merges can be computationally very costly. Our interval-based representation of melody limits the possibilities somewhat, since only states with the same pitch-label need be considered for merging. But we must still try to identify the best merges early, so that the algorithm does not waste time considering poor choices while the graph is large.

One efficient search strategy is to select at each step the state-merge that results in the lowest cost function. This procedure works well most of the time; however, the best *immediate* choice need not lead to the lowest cost *overall*. We have found it useful occasionally to "veto" a lowest-cost merge, choosing the second-lowest one instead, if the rejected merge involves members of different opening or cadential formulas. In our experience, such an intervention eventually leads to a set of lower-cost choices, and the resulting model is always easier to interpret.

Another way to make the merging more efficient is to identify all states corresponding to invariant pitches of phrase families, and to merge them before the start of the search. Such states correspond to:

- the last note of a cadential formula,
- the goal pitch of an (optional) opening formula, and
- the next-to-last stressed syllable of a cadential formula.

For the melodic family of Figure 7.3, these invariant pitches occur at the end of the phrase, and after the first and fourth dashed lines respectively.

7.4 Structure of Solutions

7.4.1 The HMM Structure

The sample model presented in this section was obtained from a corpus of fourteen syllabic chants in *Echos 1*, totaling 960 notes. The chants appear in the Sunday morning office (*Orthros*), and are recorded in the modern *Anastasimatarion* (Vallindras 1998). This corpus will suffice for the present illustration. Analysis of larger corpora has been performed and continues. While a larger corpus generally reveals more melodic possibilities, the findings reported in this section still apply.

The resulting model is so large that it cannot be usefully displayed as a graph unless we identify some meaningful substructure. The simplest way is to isolate subgraphs that correspond to a phrase's structure, as outlined in Section 7.2. (Figure 7.13).

One opening formula and five cadential formulas were identified and separated from the rest of the graph; the latter corresponds to the phrase's "body." Figure 7.12 shows the overall structure of the model in terms of the connectivity of its subgraphs. Figures 7.13 and 7.14 show the internal structure of two representative subgraphs and their connection to the rest of the graph.

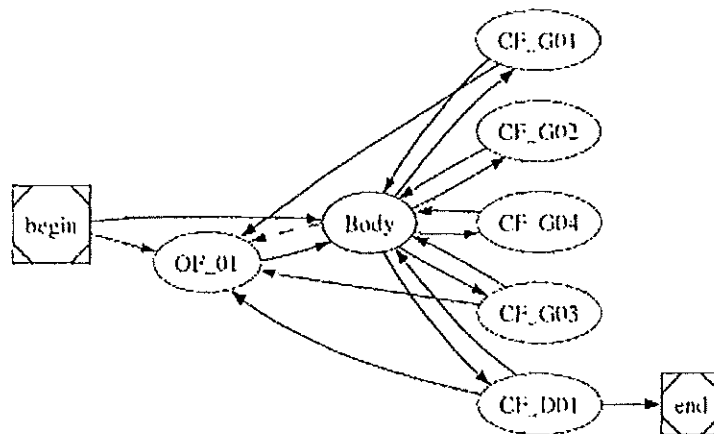


Figure 7.12. The structure of the HMM modeling Echos 1, based on a sample corpus of fourteen chants. Opening and cadential formulas are identified as subgraphs and labeled OF_x and CF_x respectively. The central part of the graph corresponds to the phrase body.

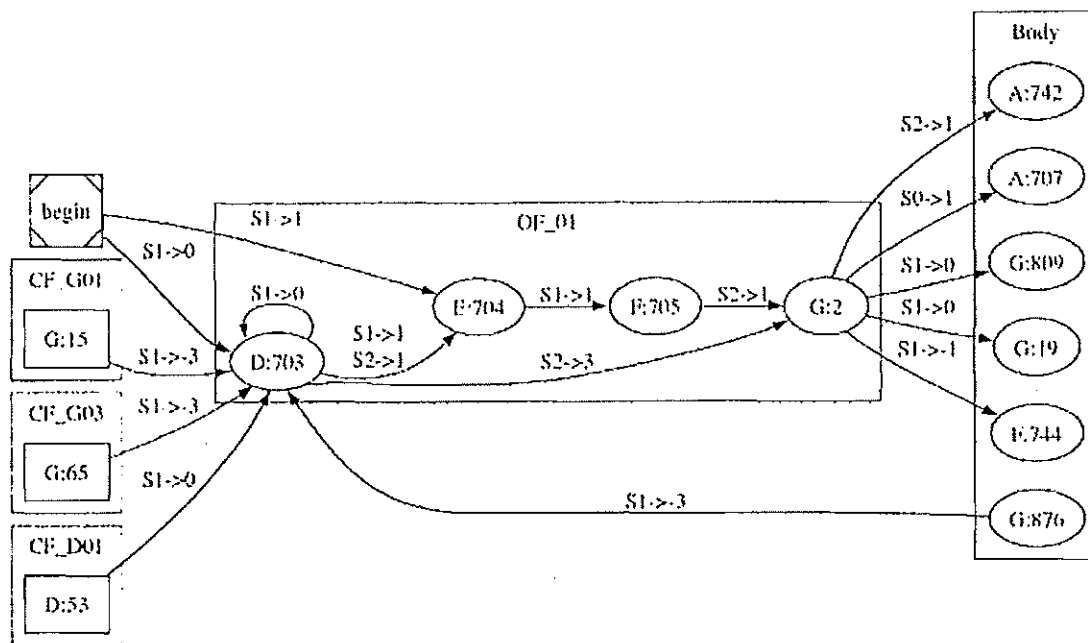


Figure 7.13. Subgraph of Figure 7.12 corresponding to the opening formula OF_{01} . States and transitions are labeled as in Figure 7.7.

7.4.2 Melodic Chunks and Decision Points

Despite the subgraph structure, the melodic properties of the *Echos* may not be entirely transparent by simply looking at the HMM graph. Moreover, we have set out to model a chanter's internalized musical knowledge, and one may question the psychological relevance of a complex representation such as ours. However, we must remember that in any given situation, a chanter need not contemplate all the melodic possibilities of the *Echos* at once. Rather, he must be able to select among those possibilities compatible with the words he has to render, typically planning and executing one phrase at a time. When we put together all possible settings of a given phrase of text, a simpler picture emerges.

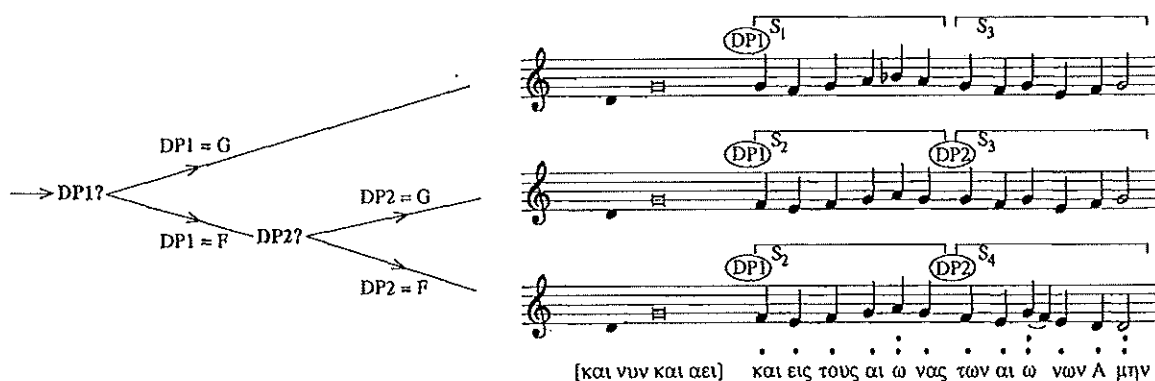


Figure 7.15. The family of all possible realizations of a given text phrase, as predicted by the HMM of Figure 7.12. The family is organized by a decision tree, with decision points marked DP1 and DP2. The melodies are thus divided up into chunks S_1 through S_4 . Chunks S_3 and S_4 are instances of cadential formulas CF_G01 and CF_D01 respectively.

Figure 7.15 shows one example, corresponding to a phrase of twelve syllables. Let us first observe that all the melodic material present in the figure derives from only four segments, which we refer to as *chunks* (they are labeled S_1 through S_4). Of these, S_3 and S_4 are realizations of cadential formulas. The remaining two chunks belong to the phrase body. In fact, S_1 and S_2 are diatonic transpositions of each other; because of their two common intervallic patterns, they both fit the stress pattern of the words. Once we know the chunk structure shared by this solution family, we can completely identify any of its members by specifying the melodic choices at only two locations, one in the beginning and one near the middle. We will refer to these locations as *decision points*. As shown in Figure 7.15, the set of melodic choices at these points can be concisely represented by a decision tree. This behavior is not unique to this particular family. Figure 7.16 shows another network of melodic possibilities corresponding to a different text phrase.

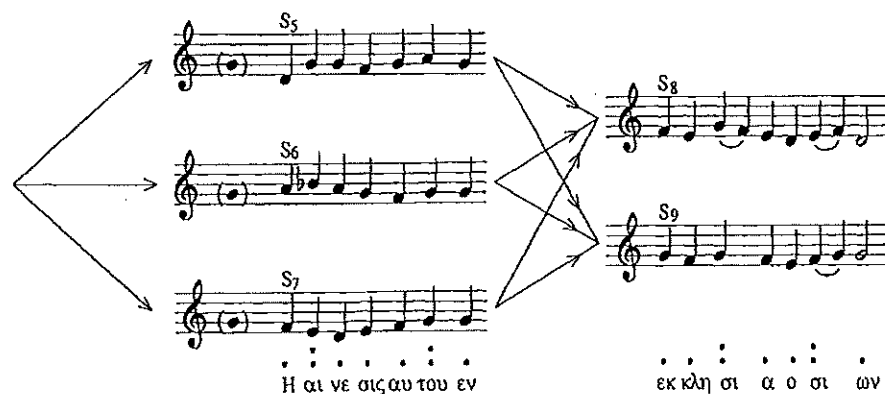


Figure 7.16. Another family of realizations corresponding to a different text phrase. The family shows an organization similar to that of Figure 7.15. Chunks S_8 and S_9 are variants of S_4 and S_3 , respectively.

Comparison of chunks occurring in other *Echoi* is currently in progress. Our results so far indicate that, even though cadential formulas are generally *Echos*-specific and fixed in pitch, opening formulas and phrase-body chunks can be diatonically transposed and can appear in several *Echoi*. This suggests the existence of stress/pitch constraints shared by all the *Echoi* that determine a chant melody's interval patterns. When combined with an *Echos*'s specific pitch system, such as the one shown in Figure 7.1, these constraints give rise to concrete melodic formulas.

7.5 Conclusions and Future Directions

7.5.1 Oral Transmission and Western Chant Studies

The question of oral transmission and how it affects melodic style has been discussed extensively in the context of Western chant scholarship. Alignment of chants similar to that of Figure 7.3 goes back to historical comparative studies of Western idioms, most notably in the work of Helmut Hucke (1955; see also Nowacki 1998). The method is often referred to as *paradigmatic* analysis. In a seminal work, Leo Treitler (1974) proposed that in order to understand the melodic idiom of Gregorian chant, one must take into account the cognitive constraints involved in oral production. In the same article, Treitler (1974: 358, 361) outlined specific models of melodic families that loosely characterize their fixed and variable parts. Edward Nowacki (1986) applied the method of melodic alignment to second-mode Old Roman tracts. He demonstrated that within these families, melodic variation involves insertions and deletions that are motivated by word stress, a result that is strikingly similar to ours. Matthew Chen (1983) studied the cadential formulas of Gregorian psalm tones. He showed how the stress pattern near the end of the verse determines the tone's cadential melody through a set of formal rules. Finally, Peter Jeffrey (1992) proposed that the question of orality in Western chant will be better addressed by studying

oral transmission in living chant cultures, in an interdisciplinary effort informed by historical musicology, ethnomusicology, and music psychology.

This work was partly inspired by all of the above studies. Following Jeffrey's program, we present an exploration of orally transmitted knowledge in the living chant tradition of the Greek Orthodox Church. We agree with Treitler's premise that the cognitive properties of melodic production will be reflected in the chant's style, and set out to capture these properties in a precise quantitative model. Following the example of Hucke, Treitler, Nowacki and others, we begin with a systematic comparison of chant melodies, aiming to uncover their underlying regularities. Borrowing techniques from computational linguistics and machine learning, we seek a model that is as complete and rigorous as that of Chen and that can fully characterize a much more complex melodic system. The psychological significance of our research remains to be addressed.

7.5.2 Towards a Cognitive Model of Melody Production

We will conclude with a discussion of our model's psychological implications, particularly in relation to chunks and decision points. We begin with decision points. We should emphasize that in our model, such points are simply formal devices that help organize a solution family; they need not be interpreted as conscious choices that take place in the course of performance. We have found, however, that once identified by a chanter, a decision point can actually assume a mnemonic role. For instance, we have seen a chanter mark a certain place in his text to remember that he must move in a given direction. Likewise, in our personal experience of practicing chant improvisation, we have found that a persistent impasse can easily be fixed once we correct a specific poorly chosen note; the rest of the melody then falls easily into place.

Let us now turn to chunks, a term which we have actually borrowed from psychology. In an influential paper, George Miller (1956) introduced the idea of chunking in the mental representation of sequences. More specifically, Miller argued that the short-term recall of a sequence deteriorates as its length increases but can be greatly facilitated if the sequence is hierarchically organized into segments, which he called chunks. Miller attributed this phenomenon to capacity limits in short-term memory; presumably a well-learned chunk is mentally encoded more compactly than the sum of its parts, and this allows for more efficient use of memory in a real-time task. A frequently cited article by Deutsch and Feroe (1981) has applied Miller's ideas directly to melody.

The chunks of our model are consistent with Miller's theory, including their typical size, which Miller estimated to be five to nine items. (In our case, it is six to nine notes.) Many controversial issues still exist concerning sequence representations: the exact chunk size and even the very notion of capacity limit have recently been

questioned (Cowan 2001). However, it seems reasonable to suppose that our chunk structure somehow facilitates real-time processing: chunks represent frequent and well-learned patterns that can be produced relatively effortlessly. As a result, only decision points require heightened attention in performance.

In this article we have shown how to construct a model of Greek chant, based on analysis of a written corpus. Melodies in that corpus have been presumably recorded—and corrected—based on a chanter's knowledge of the style, or *competence* in Chomsky's terms. To pursue the psychological implications of our model further, we eventually have to come to terms with oral outputs—including unintended ones—as they occur in real time. This is what Chomsky (1965) has described as *performance*. Putting problems of transcription aside, it is straightforward to apply our analysis to a corpus of orally produced chants. It remains to be seen whether a simple picture—one that will encompass both competence and performance while resonating with psychological theories of sequence production—will emerge.

Acknowledgements

I would like to thank David Temperley, Matthew Brown, and Gabriela Ilitchi for their feedback and encouragement during the preparation of this work.

References

- Chen, Matthew Y. (1983). "Toward a Grammar of Singing: Tune-Text Association in Gregorian Chant," *Music Perception*, 1, 84–122.
- Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.
- Chrysanthos of Madytos (1832). *Theoretikon Megates Mousikes* [Great Theoretical Treatise on Music]. Trieste: Michele Weis. Facs. edn., Athens: Ekdoseis Kouloura, 1977.
- Cowan, Nelson (2001). "The Magical Number Four in Short-Term Memory: A Reconsideration of Mental Storage Capacity" followed by Open Peer Commentary and Author's Response, *Behavioral and Brain Sciences* 24/1, 87–185.
- Desby, Frank (1974). "The Modes and Tuning in Neo-Byzantine Chant." D.M.A. Thesis, University of Southern California.
- Deutsch, Diana and John Feroe (1981). "The Internal Representation of Pitch Sequences in Tonal Music," *Psychological Review*, 88/6, 503–522.
- Giannelos, Dimitri (1996). *La musique byzantine: Le chant ecclésiastique grec, sa notation, et sa pratique actuelle*. Paris; Montréal: L'Harmattan.
- Hucke, Helmut (1955). "Gregorianischer Gesang in fränkischer und altrömischer Überlieferung," *Archiv für Musikwissenschaft*, 12, 74–87.

- Jeffrey, Peter (1992). *Re-envisioning Past Musical Cultures: Ethnomusicology in the Study of Gregorian Chant*. Chicago: University of Chicago Press.
- Karas, Simon (1981-5). *Methodos tes Hellenikes Mousikes* [Method of Greek Music]. Athens: Syllogos pros Diadosin tes Ethnikes Mousikes.
- Manning, Christopher D. and Heinrich Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press.
- Mavroeidis, Marios (1999). *Oi mousikoi tropoi sten anatolike Mesogeio* [Musical Modes in the Eastern Mediterranean]. Athens: Ekdoseis Fagotto.
- Miller, George A. (1956). "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review*, 63, 81-97.
- Nowacki, Edward (1986). "Text Declamation as a Determinant of Melodic Form in the Old Roman Eighth-Mode Tracts," *Early Music History*, 6, 193-226.
- Nowacki, Edward (1998). "Chant Research at the Turn of the Century and the Analytical Programme of Helmut Hucke," *Plainsong and Medieval Music*, 7/1, 47-71.
- Rabiner, Lawrence R. (1989). "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, 77/2, 257-286.
- Stolcke, Andreas, and Stephen Omohundro (1993). "Hidden Markov Model Induction by Bayesian Model Merging" in *Advances in Neural Information Processing Systems*, 5, 11-18, eds. S. J. Hanson, J. D. Cowan, and C. Lee Giles (San Mateo, CA: Morgan Kaufmann).
- Stolcke, Andreas and Stephen Omohundro (1994). "Best First-Model Merging for Hidden Markov Model Induction." Technical Report TR-94-003, International Computer Science Institute, University of California at Berkeley; www.icsi.berkeley.edu/techreports/1994.html (visited November 2004).
- Treitler, Leo (1974). "Homer and Gregory: The Transmission of Epic Poetry and Plainchant," *The Musical Quarterly*, 60/3, 333-372.
- Vallindras, Apostolos L., ed. (1998). *Anastasimatarion*. Athens: Ekdoseis Zoe.