Tuning a Neural Network for Harmonizing Melodies in Real-Time

Dan Gang *
Institute of Computer Science
Hebrew university
Jerusalem 91904, Israel
dang@cs.huji.ac.il

Daniel Lehmann
Institute of Computer Science
Hebrew university
Jerusalem 91904, Israel
lehmann@cs.huji.ac.il

Naftali Wagner
Department of Musicology
Hebrew university
Jerusalem, Israel
nwagner@hum.huji.ac.il

Abstract

We describe a sequential neural network for harmonizing melodies in real-time. The network models aspects of human cognition and can be used as the basis for building an interactive system that automatically generates accompaniment for simple melodies in live performance situations. The net learns relations between important notes of the melody and their harmonies and is able to produce harmonies for new melodies in real-time, that is, without advanced knowledge of the continuation of the melody. We tackle the challenge of evaluating these harmonies by applying distance functions to measure the disparity between the net's choice of a chord and that of the author of the source book from which the melody was taken. We experimented with three major issues that have implications on the performance of the model: searching for the best learning parameters (e.g., the decay parameters), the size of the learning set and the influence of metric information. The decay parameters set the scope of the short-term memory of the chords and the melody pools of units in the net. We found that the marginal benefit of a larger corpus decreases with the size of the corpus, as expected. The model contains a sub-net for meter that produces a periodic index of meter. This sub-net provides metric organization necessary for viable interpretations of functional harmonic implications of melodic pitches. We found, indeed, that representation of metric information is essential to improve the performance of harmonization as measured by the distance function.

1 Introduction

Organizing a tonal piece of music may require hyper-directionality which is a result of three parameters: functional harmony, melodic-harmonic relations and hierarchical metric structure. This work deals with neural network modeling of the relations between: harmony, melody and meter. Specifically, we describe a neural network whose task is to learn harmonization from examples in real-time situations. The system efficiently exploits the available sequential information. The net learns relations between important notes of the melody and their harmonies and is able to produce harmonies for new melodies in real-time, that is, without advanced knowledge of the continuation of the melody. Our net is fed a melody, one note at a time along with metric information. In this way the melodic information of the first beat of a measure will

*Dan Gang is supported by an Eshkol Fellowship of the Israel Ministry of Science

influence its harmony but this harmony will not be influenced by the rest of the measure's melody. This model can be used as the basis for building an interactive system that automatically generates accompaniment for simple melodies in live performance situations.

In this paper we suggest a model that is capable of learning sequences of harmonized melodies. In the generalization phase the net applies what it has learned and abstracted by harmonizing an unfamiliar melody in real-time. We tackle the challenge of evaluating these harmonies by applying distance functions to measure the disparity between the net's choice of a chord and that of the author of the source book from which the melody was taken. We also experiment with three major issues that have implications on the performance of the model: searching for the best learning parameters, the size of the learning set and the influence of metric information.

2 Corpus

The corpus contains sixty eight stylized popular western diatonic melodies. The harmonized melodies share a single meter (4/4), a single mode (major), identical lengths (16 measures), identical key (C major) and common structures.

The corpus is characterized by the simplicity of the almost pure diatonic harmonized melodies. Often, the melodies are divided into clear cut four measures phrases or sub-phrases. The range of absolute pitches of the songs is very restricted, from G4 up to C6 (with one exception E6). The table below shows the ambiti and ranges found in the learning set. Indeed, an ambitus larger than a ninth is very rare.

ambitus	*number of songs*	*range of absolute pitches*
5	8	C5-G5
6	16	G4-E5; C5-A5
7	2	B4-A5
8	21	G4-G5; C5-C6
9	10	G4-A5
10	2	A4-C6
11	2	G4-C6

The rhythmic patterns are usually very simple and repetitive. Syncopation is quite rare. The range of rhythmic values is very restricted, two up to four rhythmic values are used in one song. The shortest duration is an eighth note. Each song is divided into four hyper-measures, while significant melodic units contain one or two hyper-measures. Only fifteen different chords are used to harmonize the melodies. The chords are: the seven triads diatonically formed within the major scale, and associated dominant 7th chords for each degree (excepting the seventh scale degree which is extended by a half-diminished 7th chord with the same root) and the minor forth degree. The distribution in percentage of the fifteen chords describe in the following:

C	Dm	Em	F	G	Am	Bdim	C7	D7	E7	F7	G7	A7	Bm7b5	Fm
54.4	4.2	0.3	9.2	3	3	0	1.6	0.6	0.6	0	21.7	0.4	0	0.4

The harmonic progressions are usually simple and contain common chords in functional relationships with clearly demarcated cadences. The harmonic, melodic and metric parameters present a high level of concurrence, this is reflected for example, by the fact that on strong beats we usually find a melody note that is interpreted as a chordal note. In all songs one may find rhythmic, melodic and harmonic repetitive patterns that result in typical forms such as: AABA or AA'BA'.

3 Description of the Model

We built a sequential neural network [Jor86] that models aspects of the listening activity at the cognitive level. The model efficiently exploits available sequential information. The net learns relations between important notes of the melody and their harmonies and is able to produce harmonies for new melodies in real-time, that is, without knowledge of the melodic continuation. Other sequential neural network models that explore cognitive implication of the relations between harmony and meter can be found in [BG97] and in [BG98]. The neural network model is fed with a musical score, therefore we can assume that the "ideal" player plays the melodies' notes accurately in time and duration.

The 3-layer sequential net learns the sequence of chords, as a function of the melody's notes and the metric index. The output layer contains fifteen units, one unit for each chord, and represents the predictions or expectations of the net for the next chord. The output vector contains fifteen values of real numbers between zero and one which are interpreted as the strength of the expectations for next chord. The target chords, the melody and meter information are encoded by binary orthogonal vectors. The output layer is fed back in the state units of the input layer. The state units with the same fifteen chords represent the context of the current chord sequence. The net also includes one internal hidden layer. This hidden layer represents the chromatic scale by twelve units. The hidden layer is partially connected with the output layer, establishing the appropriate pitch to chords relations. These connections are fixed, i.e., they do not learn. The input layer contains four pools of units which are connected to the hidden layer. The first pool contains fifteen units for the state units. The second pool is the output layer of the sub-net for meter and it contains two or six units. The two units, respectively, represent the first and the third beats of the measure. The six units represent more global hierarchical metric information, such as the measure number in a musical phrase. The third pool contains twelve units representing melodic pitch classes. This pool of units is fully connected to the hidden layer, but some of the connections are fixed and some are learnable. In this way, we are able to impose external representation on the internal hidden layer. The fourth pool is plan units which are used to label different sets of notes' sequences. The chord and melody pool of units are both able to memorize context, using decay parameters that influence the scope of the context, as explained in section 4.1.

4 In Search for the Best Learning Parameters

We search for the optimal setting for decay parameters. The decay parameters set the scope of the short-term memory of the chords and the melody pools of units in the net. By so doing, we examine the idea of a flexible 'contextual window' that a performer (and listener) creates in order to optimally formulate strategies for continuation (or, in case of a listener, to build expectations for how the composer will continue). Because the number of hidden units are fixed to twelve, we experiment only with two free variables: the decay parameter of the chords and melody pools of units. It is important to note here that the whole performance of the net depends on the small initial values of the weights that are randomly chosen in the beginning of each new learning phase. Thus, to make statistical decisions about the performance of the net as a function of its decay parameters we have to obtain many samples, in hope that the samples are properly chosen so as to represent the population sufficiently well. Each sample of a population is a result of running the net for pair of specific decay parameters and with random initial weights. Each sample has three values resulting from the calculation of three simple distance functions that estimating the distance between the net's results and the original suggested by the source. In this case the probability distribution f(x) of the population is not known precisely. We take random samples (more than 30 [1]) from the population to obtain values (sample statistic) which serve to estimate the population parameters (i.e., sample mean, variance and standard deviation). On the basis of the sample information it is possible to infer statistically and to test hypotheses and significance.

4.1 Wide Search for Decay Parameters

Initially we experimented with a large range of coupled decay parameters for the chords and melodic pool of units in order to find optimal settings for decay parameters. For each couple of decay parameters we run thirty two experiments and then calculate the mean, variance and standard deviation of the sampled information. The values of the couple of the decay parameters characterized as followed: balanced couple (such as: 0.5-0.5; 0.6-0.4; 0.4-0.6), unbalanced couple (such as: 0.7-0.2) and extreme values.

The decay parameters are between 0 and 1. A large value (close to 1) means that long-past chords or notes from the melody still strongly influence the prediction of the next chord. A value close to 0 means the

[1] It is generally accepted that the mean of samples of size larger than 30 may be, for all practical purposes, assumed to be normally distributed.

memory of chords or melody is short-lived. This is because of the update rule of the activation of the chords and melody pools, which is described here:

1. Update rule for the chords pool of units:
   Activation of chords in time t = (Activation of chords in time t-1) * (Decay parameter of chords) + (Actual activation of output in time t)

2. Update rule for the melody pool of units:
   Activation of melody in time t = (Activation of melody in time t-1) * (Decay parameter of melody) + (External activation of the new melody note of time t)

4.2 Distance Functions

The quality of the harmonization of the net is estimated by three simple distance functions. For each sample the errors or the success are weighted and summed by comparing the actual output vectors with the target vector taken from the original harmonization suggested by the source.

The distance function are:

- Error function - sum of the square of the differences between output and target divided by the number of output (or target) vectors.
- Success function - sum of each match of the maximum of output with the target
- A priori weighted success function - weighted sum of each match of the maximum of output with the target. This is a payoff function that weighted the contribution of each match by multiplying with one minus the probability distribution of the chord in the corpus (see distribution of the chords in section 2). Behind this formulation is the assumption that the guess of a common chord (for example, the tonic) is providing less information than a guess of a rare chord. Therefore, the weighted function giving higher value for the guess of the rare chord than for the tonic.

The role of the three distance functions is to estimate the distance between the harmonization proceed by the system and the harmonization in the source book. None of the three distance functions takes into consideration the aesthetic aspects of the results. The distance functions do not add points for a near miss (such as predicting a chord which is an equivalent substitution), or the opposite way - do not punish in a case the chord produced is functionally or esthetically unacceptable. Dynamic changes of the context (e.g., the omission of the dominant seventh leading to the final tonic, when approaching the last measures of a song, is more problematic than in other metric locations) are not taken into account. No compensation is performed in a situation in which harmonic correction appears in a shifted metric location. As a consequence, no aesthetic claim is made here. Building a good distance function that correlates with aesthetic judgment is an extremely complex task and seems to us as a try to approach algorithmically to the harmonization problem. Moreover, we wanted to keep the formulation of the distance function clean and simple. In general we prefer the use of the third distance function. This preference is adequate for such a corpus, where the guess of just one monotonic C chord result in a hit of more than fifty percents. Following are some representative results calculated with the third distance function. Thirty two trials were performed for each couple of decay parameters and the average and the standard deviation are presented here:

Decay of Chord	*Decay of Melody*	*Average*	*Standard deviation*
0.5	0.6	69.9013	1.2426
0.4	0.4	68.0627	0.9784
0.3	0.9	59.9428	0.7409
0.95	0.5	66.3790	1.3643

4.3 The Optimal Values for the Decay Parameters

Two more experiments were performed to check the values of the decay parameters, the soundness of the architecture and the representation. The net was trained without chord context (0.0-0.5) and without chord and melody context (0.0-0.0). In this last case the only input the system has is an external melody note for each time step. One can see from the following table that when chord and melody contexts are provided, the results obtained by the third distance function are far better.

Decay of Chord	*Decay of Melody*	*Average*	*Standard deviation*
0.0	0.5	66.4771	0.9319
0.0	0.0	65.3709	1.2934
0.7	0.5	73.4633	1.0818

A global optimum value is found in 0.7 and 0.5 for the couple of decay parameters of the pool of chords and the pool of melody, respectively. We search nearby values to find what are the exact optimal values for the decay parameters. For each couple of decay parameters the performance of the net, as measured by the third distance function, is a random variable (due to different initial conditions). Each of these random variables sampled thirty two times, and we computed the: average, variance and standard deviation. Again, the values 0.7-0.5 result as the optimum decay values. Student's t test was used to compare the means of the different random variables([Leh64]). The test decides whether to accept the null hypothesis, or to reject it and accept the alternative hypothesis. The null hypothesis is that the mean estimated by the sample mean obtained for the optimum, is equal to the mean for another pair of decay values. Acceptance (or rejection) of the hypotheses is found at various levels of significance. We have thirty two samples, so the number of degrees of freedom is equal to sixty two (n1 + n2 - 2 = 32 + 32 - 2 = 62). For these degrees of freedom, we reject the null hypotheses at a 0.05 level of significance, if T is greater than 1.67. The results are summarized in the following table:

Decay	*Mean*	*Var*	*T*	*Decay*	*Mean*	*Var*	*T*	*Decay*	*Mean*	*Var*	*T*
0.6-0.4	68.49	41.10	3.17	0.6-0.45	68.84	51.05	2.77	0.6-0.5	70.05	52.25	2.03
0.6-0.55	71.05	48.40	1.46	0.6-0.6	67.60	29.59	4.04	0.65-0.4	70.14	27.98	2.32
0.65-0.45	71.53	51.67	1.15	0.65-0.5	69.09	33.90	2.92	0.65-0.55	68.98	34.05	2.99
0.65-0.6	69.88	58.38	2.06	0.7-0.4	70.26	65.72	1.78	0.7-0.45	72.41	52.48	0.62
0.7-0.5	73.46	37.45	0.00	0.7-0.55	68.96	39.00	2.90	0.7-0.6	71.91	30.81	1.06
0.75-0.4	70.31	49.16	1.91	0.75-0.45	69.66	54.10	2.24	0.75-0.5	70.19	46.57	2.01
0.75-0.55	69.62	35.05	2.55	0.75-0.6	69.94	30.36	2.41	0.8-0.4	70.72	70.63	1.49
0.8-0.45	70.17	47.36	2.02	0.8-0.5	69.19	50.24	2.58	0.8-0.55	71.61	49.84	1.12
0.8-0.6	68.96	33.50	3.02	- - - - -	- -	- -	- -	- - - - -	- -	- -	- -

For sixty degrees of freedom, the Student's t distribution is:

Tp:	55%	60%	70%	75%	80%	90%	95%	97.5%	99%	99.5%
T value:	.126	.254	.527	.679	.848	1.30	1.67	2.00	2.39	2.66

From the results presented above, one sees that for some values of decay parameters we can not reject the null hypotheses (e.g., for 0.7-.0.45 the T value is : 0.62 the significance level is between 70% up to 75%). In another experiment 900 samples were produced for decay values 0.7-0.5 and for 0.7-0.45. Each of the 900 samples were divided into thirty groups of thirty trials, in this way we deal with thirty means of the means of thirty trials. We found the following:

Decay of Chord	*Decay of Melody*	*Average*	*Standard deviation*
0.7	0.5	71.1161	0.2237
0.7	0.45	71.0544	0.1963

T value is equal to 0.20739 which is far from significance level for rejecting the null hypotheses. From the above results we conclude that the optimum value is found in the range of 0.65 up to 0.7 for the decay of the pool of chord units and 0.45 up to 0.5 for the decay of the pool of melody units.

5 The Influence of metric Information

The neural network model contains a sub-net for meter that produces a periodic index of meter. This sub-net provides metric organization necessary for viable interpretation of functional harmonic implications of melodic pitches. This section describes experiments on various possible representations of meter.

First, we experimented with no metric information at all (zero units) results in low performance as estimated by the distance function, with mean of 62.51 and standard deviation of 0.71. Then we experimented with representation of the metric information of the first beat and the third beat encoded with two units by orthogonal vectors (i.e., 10 and 01). This representation is a local metric representation and does not take into consideration hierarchical metric relations. The results are presented in the table of sub-section 4.3. We can conclude from this that the metric information is essential to improve the performance of harmonization as measured by the distance function.

For the last metric representation, the learning phase is repeated for thirty two trials. For these trials the net produces thirty two sets of different harmonizations, each set contains harmonization of the same seven unfamiliar songs. We chose three sets of harmonizations and the third author examined the results. He was able to identify exactly the three groups of harmonizations. Our conclusion is that different trials produce nets that have learned different principles, or different styles. Each net tend to produce similar harmonizations and similar harmonization errors on the different unfamiliar songs.

We tried three more global representations of the metric information using six units for the metric sub-net: from the point of view of the distance function, the result obtained were similar to those of sub-section 4.3. Detailed musical results from one of those global representations may be found in section 8.

6 Using Different Generalization Sets

All the experiments presented up to here were a result of learning of the same sixty one examples and the generalization of the same set of seven songs, chosen randomly. Nevertheless, we do not have an idea if this set is simpler to harmonize or much harder than the other possible sets. The following experiment answers this question. Thirty different sets of seven songs were randomly chosen. For each choice thirty two trials were performed and their means were calculated. The general performance as evaluated by the third distance function is: average of the thirty means is 77.88079 and the standard deviation of the average of the means is 8.2646. Our conclusion is that the previous choice (mean: 73.4633) fits well in the average range.

7 The Size of the Corpus

In a previous work([GLW97])we described a learning set that contained eighteen examples and four more examples for the generalization phase. We expand the corpus to contain sixty eight popular diatonic melodies, seven of them are randomly chosen and kept for the generalization phase. All the results presented until now results from learning sixty one examples and using seven songs for generalization. We check in this experiment if the quality of performance is affected by the number of examples in the learning set, and if yes in what way. We conducted series of experiments similar to the experiment described in section 6. But here, we randomly chose thirty times a number of learning examples and seven generalization examples. Then experiment thirty two times for each set from the thirty, and calculate their mean and then the average of the thirty means. This procedure is repeated for larger and larger number of examples for the learning set. The results are summarized in the following table:

number of learning examples	*mean and standard deviation*	
1	62.5816	1.4505
2	66.1319	1.6122
5	70.7384	1.3371
10	69.4611	1.8146
15	75.3572	1.6359
20	75.6607	1.6545
40	76.2242	1.9772
61	77.8807	1.5089

We used Student's t test for the means of sixty one learning examples and fifteen to decide if it is significant to enlarge the number of the examples in the learning set. We found that T is equal to 1.1338 with 58 degrees of freedom, which is between 80% up to 90%, so we can not reject the null hypothesis. We find that the marginal benefit of a larger corpus decreases with the size of the corpus, as expected. The experiments performed do not enable us to conclude that there is a significant benefit in using a corpus larger than fifteen.

8 Musical Results

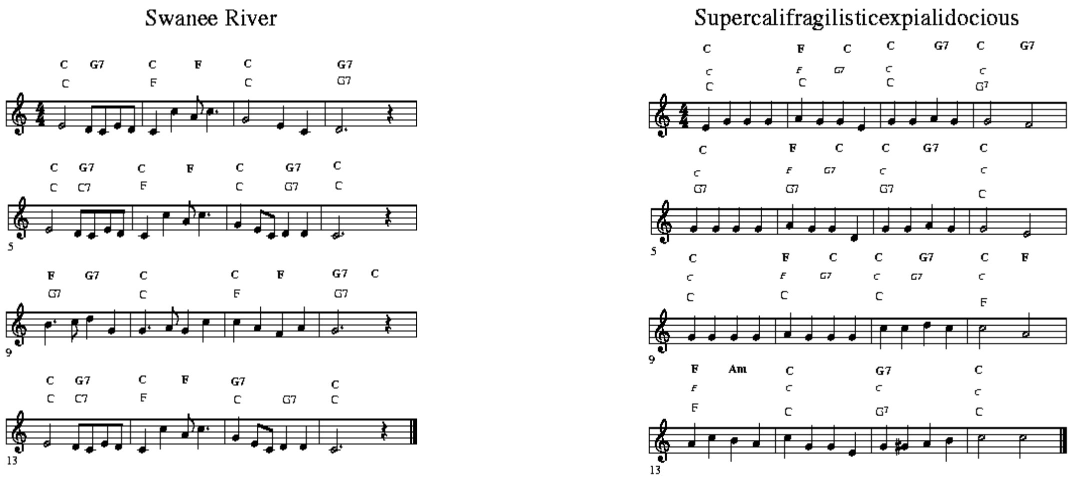

Figure 1: *Harmonization of two songs: for each song, the upper harmonization is the output of the neural network system, the lower one is found in the book. In the right side of the figure the middle harmonization was obtained from another real-time system from a previous work.*

The network's generalization capability has been tested by giving it new melodies to harmonize. In this section we present results that are produced for the decay parameters: 0.7-0.5 and for one of the more global representations alluded to in section 5. The third distance function measured the distance of the harmonization of the seven songs in the generalization set as 82.428. We present here two examples and point on some typical patterns and analyze the results in light of many examples we examined. In order to describe the results we adopt the following terminology:

- Concurrent harmonization - a melodic pitch is interpreted as a harmonic pitch class.
- Non-concurrent harmonization - a melodic pitch is interpreted as a non-chord tone.
- The concurrent harmonization or the non-concurrent harmonization is checked only for locations of beat one or beat three.
- We also use the notation MxBy (where x is between 1 to 16 and y is 1 or 3) to mark the location of measure number x in beat y.

The chords resulting from the harmonization of the song *Swanee River* (see Figure 1) is found to be functionally quite appropriate by trained musicians, if we take into consideration the real-time constraints. A typical error in real-time is concurrent harmonization of non-concurrent event in the original harmonization. Such an example is found in M1B3 and M5B3. The C chord of M2B1 is a musically required consequence for the G7 chord that is provided by the net in M1B3. Non-concurrent harmonization such as of M9B1, while the original harmonization is concurrent, are quite rare. A typical frequent error is the harmonization of *C* note with C chord instead of F chord. In all these cases the chord that rectified the error, appears with a delay of a half measure. Such errors seem unavoidable in real-time harmonization and the delayed correction produced by the system seems acceptable: additional melodic hints are accumulated during the continuation of the melody in the measure and influence the harmonic interpretation, while at the beginning of the measure this information is not available. It is no wonder the C7 chord was not predicted by the net in M5B3. Secondary dominants are used very rarely (see section 2) and they create events that require a resolution in future. More than that, notice that the original harmonization in M5B3 is non-concurrent.

The chords resulting from the net's harmonization for the song *Supercalifragilisticexpialidocious* (see Figure 1) are functionally quite appropriate. Although the net learns regularities and generalizes according to prototypes patterns, which means frequently using concurrent harmonization, the net is able to produce non-concurrent harmonization. Such examples are: the G7 chord in M3B3, the Am chord in M13B3 and G7 in M15B3. Only the last case example matches the original harmonization. In spite of the high probability of harmonizing a *G* note by C chord, the net tends to harmonize the *G* note with a G7 chord when approaching to the final tonic. This behavior, learned from the examples, improves considerably the performance of the net. The middle harmonization presents results of the work described in [GLW97]. There we claimed that: "The fact that the net is able to use only the melody's notes on the first and third beats for each measure, may lead to the wrong choice of a G7 chord for the second half of measure 2 and the lack of G7 on measure 15. In these two cases the information of the notes in the fourth beat, that is not available in real-time, might help in choosing the right chords for beat three. ... However, this does not explain why the net chose a C chord for the second half of measure 4. The problem might be the lack of accumulation of information from the beginning of the measure. This problem could be, perhaps, cured by memorizing some of the melody context". This work demonstrates that, indeed, the introduction of a melodic context, together with a finer tuning of the decay parameters, a more precise representation of metric information and an extended corpus provide for improved results.

9 Future Work and Acknowledgments

In this work, we compared the performance of networks with different parameters. A very intriguing question is: how well do our networks perform compared to musicians faced with the same real-time task? A test with human subjects is planned in the near future.

In this work we evaluated the results by measuring in an unsophisticated way the distance between the book's and the net's harmonizations. Such an evaluation does not allow for aesthetic criteria. We are looking into ways of obtaining aesthetic judgments from musicians comparing the harmonizations produced by the net and by musicians.

In addition to the practical benefits of automated accompaniment, this model may contribute to our understanding of the cognitive aspects of the harmonic and melodic inferences performed by a listener. Real-time accompaniment and expectations of a listener share a number of tasks including: correlating sequential and temporal data, memorization and contextualization of past events for the purpose of predicting the next sequential element, awareness of the location in the metric hierarchy and structure, harmonic and melodic and metric expectations and hierarchical reductive processes. While modeling cognition remains theoretical and often relies upon hypothetical interpretation, the task of real-time harmonization allows us to evaluate system performance in very specific and pragmatic terms.

We want to thank Ran El-Yaniv for helping us try to get the statistics right and to Jonathan Berger for his comments.

References

[BG97] J. Berger and D. Gang. A neural network model of metric perception and cognition in the audition of functional tonal music. In *Proceedings of the International Computer Music Association*, Thessaloniki, Greece, 1997.

[BG98] J. Berger and D. Gang. A computational model of meter cognition during the audition of functional tonal music: Modeling a-priori bias in meter cognition. In *Proceedings of the International Computer Music Association*, Michigan, 1998.

[GLW97] D. Gang, D. Lehmann, and N. Wagner. Harmonizing melodies in real-time: the connectionist approach. In *Proceedings of the International Computer Music Association*, Thessaloniki, Greece, 1997.

[Jor86] M.I. Jordan. Attractor dynamics and parallelism in a connectionist sequential machine. In *Proceedings of The Eighth Annual Conference of the Cognitive Science Society*, Hillsdale, N.J., 1986.

[Leh64] E. L. Lehmann. *Testing Statistical Hypotheses*. A Wiley publication in Mathematical statistics, 1964.