UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT

The equivalence of finite valued transducers (on HDTOL languages) is decidable

Karel Culik II
Juhani Karhumäki

Research Report CS-86-01

January 1986

The equivalence of finite valued transducers (on HDTOL languages) is decidable *

Karel Culik II

Department of Computer Science

University of Waterloo

Waterloo, Ontario, Canada

N2L 3G1

Juhani Karhumäki †

Department of Mathematics

University of Turku

Turku, Finland

Research Report CS-86-01

* This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant A-7403.

† This work has been done during the second author's visit at the University of Waterloo.

ABSTRACT

We show a generalization of the Ehrenfeucht Conjecture: For every language there exists a (finite) test set with respect to normalized k-valued finite transducers with bounded number of states. Further, we show that for each HDTOL language such a test set can be found effectively. As a corollary we solve an open problem by Gurari and Ibarra: The equivalence problem for finite valued finite transducers is decidable. This is the first time the equivalence problem is shown to be decidable for a larger class of multivalued transducers.

1. Introduction

Let L be an arbitrary language over Σ. We say that a finite subset F of L is a **test set** (with respect to morphisms) for L if whenever two morphisms agree on F they agree on L as well. The **Ehrenfeucht conjecture**, cf. [CS2] and [K], states that each language possesses a test set. It has been shown valid recently by Albert and Lawrence [AL1], and independently by Guba [Gub]. In [McN,P,Sa,St] several variations of the latter proof have been given, while in [AL2] the conjecture has been generalized for k-bounded substitutions, i.e., for substitutions satisfying that the cardinalities of the images of letters are bounded by a fixed constant k.

We show here that the Ehrenfeucht conjecture holds also for much more general types of mappings, namely for those defined by normalized finite transducers with bounded number of states and bounded degree of nondeterminism on inputs of length one. (The term "normalized" refers to the fact that the transducer reads in a single step either the empty word or a letter.) In particular, it follows that the conjecture holds for normalized k-valued transducers with bounded number of states. Moreover, we show that for this family a test set can be effectively found for each HDTOL language. Our main motivation is a corollary of this result, the decidability of the equivalence problem for k-valued finite transducers. In order to put our theorem into perspective we next list the history of main results on the equivalence of finite transducers.

Equivalence problems for various types of finite transducers (finite automata with outputs) have been extensively studied since the beginning of automata theory. In his famous paper [Mo] Moore showed that the equivalence problem is decidable for length preserving deterministic sequential machines. This decidability result was then gradually extended as follows: For deterministic *gsm*'s Jones and Laaser [JL], cf. also Blattner and Head [BH2], for single-valued transducers Schützenberger [Sc] and independently Blattner and Head [BH1], for deterministic two-way transducers Gurari [Gur1] and [Gur2], and finally for single-valued two-way transducers the authors [CK3]. A strictly larger class than that of deterministic *gsm*'s, but incomparable with the other classes above, is the class of deterministic two-tape acceptors, for which the decidability of the equivalence problem was proved by Bird [Bi].

On the other hand, Fischer and Rosenberg [FR] showed the undecidability of the equivalence problem for nondeterministic finite transducers. At the same time Griffiths [Gr] generalized this undecidability for ϵ-free nondeterministic *gsm*'s, and finally Ibarra [I1] proved it for ϵ-free *gsm*'s with unary input (or output) alphabet. Here, we further narrow the gap between the decidable and undecidable equivalence problems for finite transducers. We show that the problem is decidable for finite valued finite transducers. We give a detailed proof for the case of one-way finite transducers, but using the techniques from our previous paper [CK3] the result can be straightforwardly extended to finite valued two-way finite transducers, too.

Actually, we prove a considerably stronger result, namely that for every HDTOL language and two natural members n and k there effectively exists a test set with respect to normalized k-valued finite transducers with at most n states. Clearly, this result implies that given an HDTOL language L and two finite valued transducers we can test whether they are equivalent on L.

The decidability of testing the equivalence of mappings of certain types on languages from a family **L** has been considered by many authors. Most relevant results from the point of view of this paper can be listed as follows.

Testing the equivalence of mappings which are realized by finite transducers on a regular language is a special case of the equivalence problem for finite transducers, since a restriction of a finite transduction to a regular set is again a finite transduction. In [KK] it was shown that this problem is decidable for (multivalued) mappings of the form "morphism followed by inverse morphism", while it is undecidable for the mappings of the form "inverse morphism followed by morphism". This latter undecidability result was generalized in [Ma] for inverses of finite substitutions. A lot of attention was given to the problem of deciding the equivalence of two morphisms on a given language. Indeed, in order to prove the decidability of the DOL sequence equivalence problem in [CF] it was shown that the equivalence of two morphisms can be tested on the DOL language generated by one of them. Subsequently the following cases were shown to be decidable. Morphisms on a context-free language in [CS1], morphisms on an HDTOL language in [CK2], and as generalizations of [CS1] and [CK2] single-valued finite transducers on a context-free language in [C], and single-valued two-way finite transducers on an HDTOL language in [CK3]. For any family **L** of languages which is effectively closed under inverse morphisms and intersections with regular sets, and such that languages in **L** have effectively constructible semilinear Parikh maps, another generalization of [CS1] was shown in [I2]: It is decidable whether two deterministic two-way finite transducers are equivalent on a language from **L**.

Some of the above decidability results are based, or can be based, on the effective existence of test sets for some families of languages, which is known to hold in the following cases: For regular languages [CS2], for context-free languages [ACK], for supports of k-rational formal power series, with k a field, [RR], for DOL languages [Ru], and for HDTOL languages [CK2]. The last result in this list was generalized in [CK3], where it was shown that each HDTOL language possesses a test set even with respect to single-valued two-way (sequential) transducers with bounded number of states. We shall see here that this result can be further extended to the corresponding family of k-valued transducers, but that it does not hold in general, or even for *gsm*'s with bounded number of states.

When considering a restricted class of transducers it is certainly important whether this class is effectively given, that is whether we can test the membership in the class for an arbitrary transducer. To test whether a given finite transducer is a (deterministic) *gsm* is trivial. The decidability of single-valuedness for one-way finite transducers was shown in [Sc], and for two-way finite transducers in [CK3]. It should be noted that in both these cases the decidability of the single-valuedness actually easily implies the decidability of the equivalence problem. In [GI] it has been shown decidable whether a given finite transducer is k-valued for a given k. As will be outlined in Section 6, our earlier result to decide the single-valuedness for a two-way finite transducer can be extended to decide the k-valuedness as well.

The rest of this paper is organized as follows. In Section 2 we review some basic notions in order to fix our terminology and establish an important generalization of the Ehrenfeucht conjecture in terms of systems of equations. In the next section we introduce transducer schemata as an auxiliary device in proving the existence of a test set with respect to quite a large family of transducers. It is shown that such a test set exists (noneffectively) for each language. In Section 4 we show that for each HDTOL language a test set with respect to normalized

k-valued transducers with bounded number of states exists effectively. Since regular sets are a subfamily of HDTOL languages, this result implies the decidability of the equivalence problem for k-valued transducers. In Section 5 we show that for a given regular R and $n \geq 1$ a test set with respect to all normalized finite transducers (or even *gsm*'s) having at most n states cannot be found effectively. In the last section we introduce the notion of a semideterministic (having deterministic transitions but possibly nondeterministic outputs) finite transducer and show that their equivalence problem is decidable if and only if the problem of testing the equivalence of two finite substitutions on regular languages is decidable, cf. [CK4]. Finally, we discuss the generalizations of our main results to two-way k-valued finite transducers.

2. Preliminaries and an auxiliary result

We assume that the reader is familiar with the basic results on formal languages [H], finite transducers [Be] and L systems [RS]. Consequently, the following lines are mainly to fix our terminology.

A **finite transducer** is a sixtuple $T = <Q, \Sigma, \Delta, s_0, F, E>$, where Q is a set of states, Σ and Δ are input and output alphabets, respectively, s_0 is an initial state, F is a set of final states and $E \subseteq Q \times \Sigma^* \times \Delta^* \times Q$ is a finite set of transitions. We write $q \xrightarrow{u,\, v} p$ if T goes from state q into state p by reading input u and producing output v. Transducer T realizes a **finite transduction** $|T| : \Sigma^* \rightarrow \Delta^*$ or rational **relation** $R_T \subseteq \Sigma^* \times \Delta^*$ as follows: $y \in |T|\, x$ (or $(x, y) \in R_T$) if $s_0 \xrightarrow{x,\, y} p$ for some $p \in F$. It is well known that every rational relation can be realized by a **normalized** finite transducer i.e., with $E \subseteq Q \times (\Sigma \cup \{\epsilon\}) \times \Delta^* \times Q$. (In fact, also by a transducer with $E \subseteq Q \times (\Sigma \cup \{\epsilon\}) \times \Delta \cup \{\epsilon\} \times Q$). Finally, a finite transducer is called a

generalized sequential machine (*gsm*) if $E \subseteq Q \times \Sigma \times \Delta^* \times Q$.

We say that a finite transducer T admits an (input) ϵ-loop if there is $q \in Q$, $v \in \Delta^+$ such that $q \xrightarrow{\epsilon,\, v} q$. If T does not admit an ϵ-loop it is **ϵ-loop-free**. Clearly, for each ϵ-loop-free finite transducer T the cardinality of $|T|\, x$ for any x in Σ^* is finite. Let k be an nonnegative integer. We say that a transducer T is **k-valued** if, for each x in Σ^*, $|T|x$ contains at most k words, and that it is **finite valued** if it k-valued for some k. Two finite transducers T_1 and T_2 are **equivalent on a language** L if, for each word w in L, $|T_1|\, w = |T_2|\, w$. In particular, if this holds for the language Σ^* (or equivalently for $dom(T_1) \cup dom(T_2)$) we say that T_1 and T_2 are **equivalent**.

We conclude this section with an important auxiliary result, essentially proved in [AL2] as a consequence of the validity of the Ehrenfeucht conjecture. For the sake of completeness we also give the proof of this result. First, however, we have to introduce some more terminology.

Let V be a finite set (of variables) disjoint from our basic alphabet Σ. An **equation** over Σ with V as a set of variables is a pair (u, v) in $V^* \times V^*$ and its **solution** is any morphism $h : V^* \to \Sigma^*$ satisfying $h(u) = h(v)$. Two systems of equations (i.e., sets of equations) are called **equivalent** if they have exactly the same solutions. As shown in [CK1] the Ehrenfeucht conjecture can be stated as follows: Any system of equations with a finite set of variables is equivalent to its finite subsystem. The set of natural numbers is denoted by $\mathbb{N}$. Let $k : \mathbb{N} \to \mathbb{N}$ be a function and let, for each pair (i, j) of natural numbers satisfying $1 \leq j \leq k(i)$, $S(i, j)$ be a finite system of equations over Σ with the same set V of variables. We consider the formula

$$\mathbf{S} = \bigwedge_{i=1}^{\infty} \bigvee_{j=1}^{k(i)} S(i, j)$$

and say that a morphism $h : V^* \to \Sigma^*$ is a **solution** of **S** if, for each $i \geq 1$, there exists an index j_i, with $1 \leq j_i \leq k(i)$, such that h is a solution of $S(i, j_i)$. The claim that **S** is equivalent to its finite subpart now means, of course, that there exists an $m \geq 1$ such that each solution of the formula $\bigwedge_{i=1}^{m} \bigvee_{j=1}^{k(i)} S(i, j)$ is a solution of **S** as well. Now, we are ready for our auxiliary result:

Theorem 1. For each formula $\mathbf{S} = \bigwedge_{i=1}^{\infty} \bigvee_{j=1}^{k(i)} S(i, j)$, where k is a function from $\mathbb{N}$ into $\mathbb{N}$ and each $S(i, j)$ is a finite system of equations, there exists an $m \geq 1$ such that **S** is equivalent to the formula $\bigwedge_{i=1}^{m} \bigvee_{j=1}^{k(i)} S(i, j)$.

Proof. We associate with **S** a finitely branching infinite tree $T_{\mathbf{S}}$ labelled by finite systems of equations as follows. The root node (labelled by the empty set of equations) has descendants $S(1, j)$, $j = 1, \ldots, k(1)$, and each node on level i labelled by $\bar{S}$ has descendants $\bar{S} \cup \mathbf{S}(i+1, j)$ for $j = 1, \ldots, k(i+1)$. So each path in this tree defines an ascending chain of finite systems of equations and a morphism h is a solution of **S** iff it is a solution of (at least) one infinite system of equations defined by an infinite path in $T_{\mathbf{S}}$.

Now, we cut off the "unnecessary" branches of $T_{\mathbf{S}}$ as follows. We say that a node s in $T_{\mathbf{S}}$ is **terminal** if there exists an infinite path starting from that node such that each system of equations appearing as a label of a node in this path is equivalent to the system of equations appearing as a label of s. We throw away all descendants of terminal nodes and let a tree thus obtain be $\bar{T}_{\mathbf{S}}$. If $\bar{T}_{\mathbf{S}}$ would contain an infinite branch, then, by the construction, for each system of equations S occurring as a label of a node in this path there exists another node labelled by a system S' of equations such that S and S' are not equivalent. This, in turn, would yield an infinite system of equations that does not possess a finite equivalent subsystem. This, however, is impossible by the Ehrenfeucht conjecture, see [AL1], therefore $\overline{T}_S$ is finite.

It also follows directly from the construction that $\overline{T}_S$ is equivalent to T_S in the sense that any solution of a system of equations occurring as a label of a terminal node in $\overline{T}_S$ is a solution of a system of equations defined by an infinite path in T_S, in other words, it is a solution of **S**. So we can choose m to be the length of the longest path in $\overline{T}_S$. □

3. Existence of a test set

Let $L \subseteq \Sigma^*$ be a language and Θ a family of finite transducers having Σ and Δ as input and output alphabets, respectively. We say that a finite subset F of L is a **test set** for L with respect to Θ if, for any two transducers from Θ, they are equivalent on L if and only if they are equivalent on F. In particular, if Θ is a family of morphisms, then we have the ordinary notion of a test set, cf. [CS2], and the recently proved Ehrenfeucht conjecture states that such an F always exists, cf. [AL1] and [Gub].

Our goal here is to generalize this result for families of transducers. Clearly, a test set does not exist for every L with respect to the family of all finite transducers. On the other hand, we shall show that it exists for each L with respect to the family of ϵ-loop-free, normalized finite transducers satisfying the following two conditions: the number of states is bounded by a fixed (but arbitrary) constant, and the **degree of edge-ambiguity**, that is the maximal cardinality of the set $E \cap (\{q\} \times \{a\} \times \Delta^* \times \{q'\})$ where q and q' range over Q and a over $\Sigma \cup \{\epsilon\}$, is bounded by another fixed (but arbitrary) constant d. Let us denote this family of finite transducers by $T_{n,d}(\Sigma, \Delta)$.

Theorem 2. Let n and d be natural numbers. Each language $L \subseteq \Sigma^*$ possesses a test set F with respect to $T_{n,d}(\Sigma, \Delta)$.

Proof. In order to group $T_{n,d}(\Sigma, \Delta)$ into a finite number of cases we introduce the notion of transduce schema as in [CK3]. A **transducer schema** over a finite set Ω is a ϵ-loop-free, normalized transducer T_Ω, with Σ and Ω as input and output alphabets, and satisfying the following three properties:

(i) if $(q, a, u, q') \in E$, then $u \in \Omega$,

(ii) the edge-ambiguity of T_Ω is at most d,

(iii) if $(q, a, u, q') \in E$ and $(p, b, v, p') \in E$ with $(q, a, q') \neq (p, b, p')$ then $u \neq v$.

Clearly, for a fixed Σ and fixed number of states, there exists only a finite number of distinct transducer schemata (up to renaming of outputs). Thus, under these assumptions Ω can be assumed to be fixed, too. For our purposes it is illustrative to call Ω the set of **variables**.

Let $S_{n,d}(\Sigma, \Omega) = \{S \in T_{n,d}(\Sigma, \Omega) \mid S \text{ is a transducer schema}\}$. For a mapping $i : \Omega \rightarrow \Delta^*$ and S in $S_{n,d}(\Sigma, \Omega)$, we denote by $i(S)$ the transducer obtained from S by replacing in E each output (variable) c by $i(c)$. We say that $i(S)$ is an **interpretation** of S via i. Let $I(S)$ denote the set of all interpretations of S and let $I(S_{n,d}(\Sigma, \Omega)) = \bigcup_{S \in S_{n,d}(\Sigma, \Omega)} I(S)$. Obviously, $I(S_{n,d}(\Sigma, \Omega)) = T_{n,d}(\Sigma, \Delta)$, so that $S_{n,d}(\Sigma, \Omega)$ provides a "finite base" for the family of ϵ-loop-free, normalized finite transducers with Σ and Δ as input and output alphabets and having at most n states and edge-ambiguity at most d. Consequently, to prove the theorem it is enough to establish the following claim.

Assertion. Let S_1 and S_2 be two transducer schemata. There exists a finite subset $F(S_1, S_2)$ of L such that for any two transducers T_1 and T_2 from $I(S_1)$ and $I(S_2)$, respectively, T_1 and T_2 are equivalent on L if and only if they are equivalent on $F(S_1, S_2)$.

Proof of Assertion. If $L(S_1, S_2) = L \cap ((dom(S_1) - dom(S_2)) \cup (dom(S_2) - dom(S_1)))$ is nonempty we can choose $F(S_1, S_2)$ to be any singleton subset of $L(S_1, S_2)$. So assume that $L(S_1, S_2)$ is empty. Consider a word x in $L \cap dom(S_1)$. Since S_1 and S_2 are ϵ-loop-free the output sets $|S_1|x$ and $|S_2|x$ are finite, say

$$|S_1|x = \{y_1, \ldots, y_t\} \text{ and } |S_2|x = \{z_1, \ldots, z_r\} .$$

Let $Y = \{y_1, \ldots, y_t\}$ and $Z = \{z_1, \ldots, z_r\}$. Let $\mathbf{P}_Y$ and $\mathbf{P}_Z$ denote the sets of all partitions of Y and Z, respectively. For each pair (P_Y, P_Z) in $\mathbf{P}_Y \times \mathbf{P}_Z$, with the same cardinality, and for each bijection $\alpha : P_Y \to P_Z$ we define a finite set of equations as follows:

$$y_i = \bar{z}_j \text{ iff } \alpha[y_i] = [z_j] , \tag{1}$$

where the square brackets are used to denote the equivalence class defined by its representative and $\bar{z}_j$ denotes the barred copy of z_j. Clearly, the above procedure yields only a finite number of systems of equations with $\Omega \cup \overline{\Omega}$ as a set of variables.

Now, an important observation is that each of the systems of equations in (1) describes one of the possible ways in which two transducers T_1 from $I(S_1)$ and T_2 from $I(S_2)$ can be equivalent on x. That is to say, if T_1 is obtained from S_1 via interpretation i_1 and T_2 from S_2 via i_2, then T_1 and T_2 are equivalent on x if and only if the morphism $h : (\Omega \cup \overline{\Omega})^* \to \Delta^*$ defined by $h(w) = i_1(w)$ for w in Ω and $h(\overline{w}) = i_2(w)$ for $\overline{w}$ in $\overline{\Omega}$ is a solution of (at least) one of the equations in (1).

We carry out the above construction over all words x in $L \cap dom(S_1)$. This results an formula

$$\mathbf{S} = \bigwedge_{i=1}^{\infty} \bigvee_{j=1}^{k(i)} S(i, j) , \tag{2}$$

with $k : \mathbb{N} \rightarrow \mathbb{N}$ and each $S(i, j')$ a finite system of equations over $\Omega \cup \overline{\Omega}$, such that the solutions of this formula characterizes these pairs of transducers in $I(S_1) \times I(S_2)$ which are equivalent on L. But, by Theorem 1, (2) has an equivalent finite subsystem. Hence, the corresponding finite subset of L tests the equivalence of any pair of transducers from $I(S_1) \times I(S_2)$. So our proof for Assertion, and also for the theorem, is complete. □

We finish this section with a couple of remarks explaining our choice of family of transducers. First of all, in order to keep the set of variables finite (and have a "finite base" property) we have to bound the cardinality of the state set as well as the degree of edge-ambiguity. Because of the same reason we also have to bound uniformly the set of input words read in a single step, hence we consider only normalized transducers. Finally, the ϵ-loop-freeness is required in order to get only finite systems of equations in (1). Despite these restrictions our class $T_{n,d}(\Sigma, \Omega)$ is large enough to yield interesting corollaries, as we shall see in the following two sections.

We also want to emphasize that it is not only our proof techniques but the nature of the problem which requires both the parameters k and d. Clearly, Theorem 2 does not hold if the number of states is not bounded. The fact that for the case of nonbounded edge-ambiguity it does not hold either follows from a recently proved result [L] that even the simple regular language ab^*c does not possess a test set with respect to the family of finite substitutions.

4. Effective subcase

In this section we are looking for conditions under which the test set of Theorem 2 can be effectively found, while in the next section we show that this is not possible in general.

Let $T_n^k(\Sigma, \Delta)$ denote the family of k-valued, normalized transducers with Σ and Δ as input and output alphabets, respectively, and with at most n states. Clearly, the only possible ϵ-loops in k-valued transducers are of the form $q \xrightarrow{\epsilon,\epsilon} q$. These, however, can be eliminated without affecting the transduction realized by the transducer and the number of states of the transducer. It is also obvious that every ϵ-loop-free transducer in $T_n^k(\Sigma, \Delta)$ is also in our earlier class $T_{n,k}(\Sigma, \Delta)$. Hence, we have a corollary of Theorem 2.

Theorem 3. Let k and n be natural numbers. For each language L there exists a test set F with respect to $T_n^k(\Sigma, \Delta)$.

In order to state our main result of this section, which provides an effective subcase of Theorem 3, we have to recall some definitions on L systems, cf. [RS]. A DTOL system is a $(t + 2)$-tuple $(\Sigma, h_1, \ldots, h_t, w)$ where each $h_i : \Sigma^* \to \Sigma^*$ is a morphism and w is a word in Σ^+. It defines the language $L = \bigcup_{i=0}^{\infty} L_i$, where $L_0 = \{w\}$ and $L_{i+1} = \bigcup_{j=1}^{t} h_j(L_i)$ for $i \geq 0$. Languages thus obtained are called **DTOL languages** and languages of the form $h(L)$, where L is a DTOL language and h is a morphism, are called **HDTOL languages**.

Theorem 4. Let k and n be natural numbers. For each HDTOL language L there effectively exists a test set F with respect to $T_n^k(\Sigma, \Delta)$.

Proof. As it has been already stated at the beginning of this section it is enough to consider ϵ-loop-free transducers in $T_n^k(\Sigma, \Delta)$, or equivalently, we may assume that $T_n^k(\Sigma, \Delta)$ contains only such transducers. Now, the proof is split into two Assertions.

Assertion I. Given two languages $L_1, L_2 \subseteq \Sigma^*$, with $L_1 \subseteq L_2$, and two natural numbers n and k, it is decidable whether L_1 is a test set for L_2 with respect to $T_n^k(\Sigma, \Delta)$.

In order to prove Assertion I we recall the construction in the proof of Theorem 2. By that proof, we can associate each word with a finite number of finite systems of equations in such a way that the set of all the solutions of at least one of these systems characterizes all these pairs of transducers from $T_n^k(\Sigma, \Delta)$ which are equivalent on this given word. Consequently, since L_1 and L_2 are finite, the assertion follows from the fact that it is decidable whether two finite systems of equations have exactly the same solutions, cf. [CK1].

Assertion II. Let $h : \Sigma^* \rightarrow \Sigma^*$ be a morphism and let n and k be two natural numbers. If F' is a test set for $L' \subseteq \Sigma^*$ with respect to $T_n^k(\Sigma, \Delta)$, then $h(F')$ is a test set for $h(L')$ with respect to $T_n^k(\Sigma, \Delta)$.

In order to prove Assertion II we first conclude that for a transducer T in $T_n^k(\Sigma, \Delta)$, there exists another transducer $T(h)$ in $T_n^k(\Sigma, \Delta)$ such that

$$|T(h)\,|(w) = |T\,|(h(w)) \quad \text{for all } w \text{ in } \Sigma^* . \tag{1}$$

The construction of $T(h)$ is a routine one: The sets of initial and final states are subsets of those of T, and for each computation of the form $q \xrightarrow{h(a),u} q'$, with a in Σ, according to T, T_h contains a transition (q, a, u, q'), and no other transitions. Hence, the number of states of $T(h)$ is at most that of T. It also follows directly from (1) that $T(h)$ is k-valued.

Now, assume the contrary that Assertion II does not hold. Then there exist transducers T_1 and T_2 in $T_n^k(\Sigma, \Delta)$ such that they are equivalent on $h(F')$ but $|T_1|(h(x')) \neq |T_2|(h(x'))$ for some x' in L'. Hence the transducers $T_1(h)$ and $T_2(h)$ are equivalent on F', but not on L', a contradiction.

Now we are ready to finish the proof of Theorem 3. Let $L = g(L_\infty)$, where L_∞ is generated by a DTOL system $(\Sigma, h_1, \ldots, h_t, w)$. Define

$$L_0 = \{w\} \text{ and } L_{i+1} = \bigcup_{j=1}^{t} h_j(L_i) \cup \{w\} \text{ for } i \geq 0 . \qquad (2)$$

By Theorem 2, there exists an indexed i' such that $L_{i'}$ is a test set for L_∞, and hence also for $L_{i'+1}$, with respect to $T_n^k(\Sigma, \Delta)$. By Assertion I, we can find such an index i_0 that L_{i_0} is a test set for L_{i_0+1}. We claim that actually L_{i_0} is a test set for L_∞ as well. By Assertion II, for each $j = 1, \ldots, t$, $h_j(L_{i_0})$ is a test set for $h_j(L_{i_0}+1)$, and so, by (2) and elementary properties of test sets, L_{i_0+1} is a test set for L_{i_0+2}. So by the transitivity property of test sets L_{i_0} is a test set for L_{i_0+2} and our claim follows by induction. Finally, applying again Assertion II, we conclude that $g(L_{i_0})$ is a test set for L with respect to $T_n^k(\Sigma, \Delta)$. □

After having Theorem 4 a natural question arises: Is it decidable whether a given finite transducer is k-valued? The answser to this question is affirmative. Indeed, in [GI] it has been shown that the problem can be solved in polynomial time. An alternate algorithm for this problem would be obtained as a modification of our proof of Theorem 4, where test sets for the k-valuedness instead of those for the equivalence would be considered, cf. [CK3]. As a consequence of Theorem 4 and the above discussion we obtain the following strengthening of the DTOL sequence equivalence problem, cf. [CK3].

Theorem 5. It is decidable whether two finite valued finite transducers are equivalent on a given HDTOL language.

Proof. Let T_1 and T_2 be two finite valued transducers. As discussed above we can decide whether T_i's are k-valued for a fixed k, and hence find the smallest ℓ such that they are ℓ-valued. Then we construct for T_i's the equivalent normalized transducers $\overline{T}_i$. Clearly, we may assume that $\overline{T}_i$'s are ϵ-loop-free and so in $T_m^\ell(\Sigma, \Delta)$ where m is the maximal cardinality of the state sets of $\overline{T}_1$ and $\overline{T}_2$. Now, Theorem 5 follows directly from Theorem 4. □

As another corollary of Theorem 4 we have a solution of an open problem stated in [GI]:

Theorem 6. The equivalence problem for finite valued finite transducers is decidable.

Proof. Given two finite valued transducers T_1 and T_2 we first check whether their domains, which are regular sets, coincide. If "not" we are done: The transducers are not equivalent. If "yes" we proceed as in the proof of Theorem 5 and conclude our result from the fact that we can find a morphism g and a DTOL language L such that $dom(T_i) = g(L)$. This is, indeed, possible since $dom(T_i)$ is regular. □

As a concluding remark of this section, we want to emphasize that problems concerning k-valued, and hence also finite valued, transducers are essentially different from those of single-valued transducers. For example, the single-valuedness of a finite transducer can be relatively easily checked, as first shown in [Sc], and it follows directly from this decidability that the equivalence problem for single-valued transducers is decidable, as well. On the other hand, k-valuedness, for a fixed $k \geq 2$, is much more difficult to decide, and moreover, the decidability of this problem does not seem to give any algorithm to test the equivalence of k-valued transducers.

5. Noneffective subcase

Since the equivalence problem for finite transducers is undecidable, cf. [B], Theorem 4 does not hold for the whole family $T_{n,d}(\Sigma, \Delta)$. Indeed, the transducer $T(h)$ at the beginning of the proof of Assertion II can have a larger edge-ambiguity than T, and hence the number of variables associated with $T(h)$ is not the same as that associated with T. It is, however, interesting to observe that this is the only point where the proof for $T_{n,d}(\Sigma, \Delta)$ breaks down.

In this section we reduce further this noneffectiveness. For a regular language $R \subseteq \Sigma^*$ and for two natural numbers n and d let $GSM_{n,d}(R, \Delta)$ denote the family of all *gsm*'s with at most n states, with the edge-ambiguity at most d, with the domain equal to R, and with Δ as the output alphabet. We have:

Theorem 7. For natural numbers n and d and for a regular language R there exists a test set F for R with respect to $GSM_{n,d}(R, \Sigma)$, but it cannot be found effectively, in general.

Proof. The existence of a test set F follows as a special case from Theorem 2. The noneffectiveness, in turn, is a consequence of the fact that the equivalence problem for *gsm*'s is undecidable, cf. [Gr]. □

Actually, analysing the proof of [Gr] one observes that the parameter n, but not d, can be fixed (to be no larger than 13) without affecting the noneffectiveness of Theorem 7.

We believe that Theorem 7 is interesting in the sense that it is one of the few known results stating that something holds for regular languages, but provably noneffectively. In what follows we give another result of this nature.

Instead of considering the equivalence of two transducers on a language we now consider their inclusion, that is the problem of whether for a language L and for two transducers T_1 and T_2 the relation $|T_1|\, x \subseteq |T_2|\, x$ holds for all x in L. We define, in a natural way, the notion of an **inclusion test set** for L with respect to a family of finite transducers.

We restrict our considerations to the family of transductions realized by single-state *gsm*'s with the edge-ambiguity bounded by a fixed constant k, i.e., to **k-bounded finite substitutions**. Let us denote this family by $FS_k(\Sigma, \Delta)$ and its two elements by τ and σ. Obviously, we can describe the fact that $\tau(x) \subseteq \sigma(x)$, for a word x, by using a formula similar to that we used to describe the fact that two transducers are equal or a given word, cf. Section 3. Moreover, it is shown in [Ma] that the problem of deciding whether for a given regular language R and for two finite substitutions τ and σ the relation $\tau(x) \subseteq \sigma(x)$ holds for all x in R is undecidable. So in analogy to Theorem 7 we also have:

Theorem 8. For a natural number k and for a regular language $R \subseteq \Sigma^*$ there exists an inclusion test set F for R with respect to $FS_k(\Sigma, \Delta)$, but it cannot be found effectively.

Whether or not Theorem 8 holds for ordinary test sets remains open, cf. also Section 6.

6. Concluding remarks

We have generalized the Ehrenfeucht conjecture for mappings realized by families of transducers. We have also been able to exhibit quite large families of transducers (Theorem 4) such that every regular language possesses effectively a test set with respect to these families. On the contrary, we also show that for still larger families this effectiveness does not hold anymore (Theorem 7). The distinctions between these two families of transducers is that in the former case the mappings are of bounded nondeterminism, while in the second case the nondeterminism is unbounded. In terms of equivalence problems the above can be stated as follows: If the nondeterminism of the mappings realized by finite transducers is bounded by a fixed constant, then their equivalence is decidable, while if it is unbounded, then it is undecidable (even in the case of *gsm*'s).

One way to generalize deterministic *gsm*'s is to consider single-valued or k-valued finite transducers. As we saw, the decidability of the equivalence problem is not affected by this generalization. Another possibility is to require that a *gsm* is deterministic with respect to inputs, but may produce several outputs in each step, i.e., the set of transitions E satisfies the condition: if $(q, a, u, q') \in E$ and $(q, a, v, q'') \in E$, then $q' = q''$ (but possibly $u \neq v$). We call such *gsm*'s **semideterministic**.

It is open whether the equivalence problem for semideterministic *gsm*'s is decidable. However, we can show that this problem is equivalent to a simpler problem, which is conjectured to be decidable, cf. [CK4].

Theorem 9. The equivalence problem for semideterministic finite transducers is decidable if and only if the problem of deciding whether two finite substitutions are equivalent on a given regular language is decidable.

Proof. Clearly, the decidability of the equivalence of two semideterministic *gsm*'s implies the decidability of the other problem. To prove the converse let $T_i = (Q_i, \Sigma_i, \Delta_i, s_{0,i}, F_i, E_i)$, for $i = 1, 2$, be two semideterministic *gsm*'s. Let A_1 and A_2, respectively, be the underlying deterministic automata. We test their equivalence, and if the answer is negative we are done: T_1 and T_2 are not equivalent. If A_1 and A_2 are equivalent we define their crossproduct A, which is a deterministic automaton equivalent to both A_1 and A_2. We set L to be the set of all successful computations of A. Clearly, L can be described as a regular subset of $(Q_1 \times Q_2 \times \Sigma \times Q_1 \times Q_2)^*$. Further, we define finite substitutions $\tau_1, \tau_2 : (Q_1 \times Q_2 \times \Sigma \times Q_1 \times Q_2)^* \to \Delta^*$ by the formula

$$\tau_i(q_1, q_2, a, q_1', q_2') = \{u \mid (q_i, a, u, q_i') \in E_i\}, \quad i = 1, 2 .$$

It is a direct consequence of our construction that T_1 and T_2 are equivalent if and only if τ_1 and τ_2 are equivalent on L. Hence, our proof is complete. □

As our final remark we want to mention that combining the techniques used in this paper with our older one used in [CK3], the results of this paper can be strengthened for 2-way finite transducers as well. Without going into further details we only state the generalizations of Theorems 8 and 9 from [CK3] and believe that the interested reader can work out the details.

Theorem 10. The equivalence problem for k-valued two-way finite transducers is decidable.

Theorem 11. It is decidable whether a given two-way finite transducer is k-valued.

Observe that, as in the case of one-way finite transducers, either of Theorems 10 and 11 does not follow, at least immediately, from the other. The special cases, where $k = 1$, were proved in [CK3], while in [GI] Theorem 11 was shown for one-way finite transducers.

References

[ACK] J. Albert, K. Culik II and J. Karhumäki, Test sets for context-free languages and algebraic systems of equations, Inform. Control 52(1982) 172-186.

[AL1] M.H. Albert and J. Lawrence, A proof of Ehrenfeucht's Conjecture, Theoret. Comput. Sci. (to appear).

[AL2] M.H. Albert and J. Lawrence, Test sets for finite substitutions, manuscript (1985).

[Be] J. Berstel, Transductions and Context-free Language (Teubner, Stuttgart, 1979).

[Bi] M. Bird, The equivalence problem for deterministic two-tape automata, J. Comput. System Sci. 7(1973) 218-236.

[BH1] M. Blattner and T. Head, Single-valued a-transducers, J. Comput. System Sci. 15(1977) 310-327.

[BH2] M. Blattner and T. Head, The decidability of equivalence for deterministic finite transducers, J. Comput. System Sci. 19(1979) 45-49.

[C] K. Culik II, Some decidability results about regular and pushdown translations, Inform. Process. lett. 8(1979) 5-8.

[CF] K. Culik II and I. Fris, The decidability of the equivalence problem for DOL-systems, Inform. Control 35(1977) 20-39.

[CK1] K. Culik II and J. Karhumäki, Systems of equations over a free monoid and Ehrenfeucht's Conjecture, Discrete Math. 43(1983) 139-153.

[CK2] K. Culik II and J. Karhumäki, The decidability of the DTOL sequence equivalence problem and related decision problems, University of Waterloo, Department of Computer Science, Research Report

CS-85-05 (1985).

[CK3] K. Culik and J. Karhumäki, The equivalence problem for single-valued two-way transducers is decidable, University of Waterloo, Department of Computer Science, Research Report CS-85-24 (1985).

[CK4] K. Culik II and J. Karhumäki, Decision problems solved with the help of the Ehrenfeucht Conjecture, EATCS Bull. 27 (1985) 30-35.

[CS1] K. Culik II and A. Salomaa, On the decidability of morphic equivalence for languages, J. Comput. System Sci. 17(1978) 163-175.

[CS2] K. Culik II and A. Salomaa, Test sets and cfecking words hor homomorphism equivalence, J. Comput. System Sci. 15(1980) 375-395.

[FR] P.C. Fischer and A.L. Rosenberg, Multi-tape one-way nonwriting automata, J. Comput. System Sci. 2(1968) 88-101.

[Gub] V.S. Guba, personal communication (1985).

[Gur1] E. Gurani, The equivalence problem for deterministic two-way transducers is decidable, SIAM J. Comput. 11(1982) 448-452.

[Gur2] E. Gurani, Two-way counter machines and finite-state transducers, Intern. J. Comput. Math. 17(1985) 229-236.

[GI] E. Gurani and O. Ibarra, A note on finite-valued and finitely ambiguous transducers, Math. Systems Theory 16(1983) 61-66.

[Gr] T. Griffiths, The unsolvability of the equivalence problem for ϵ-free nondeterministic generalized machines, J. Assoc. Comput. Mach 15(1968) 409-413.

[H] M.A. Harrison, Introduction to Formal Language Theory (Addison-Wesley, Reading, MA, 1982).

[I1] O. Ibarra, The unsolvability of the equivalence problem for ϵ-free *NGSM*'s with unary input (output) alphabet and applications, SIAM J. Comput. 4(1978) 524-532.

[I2] O. Ibarra, 2DST mappings on languages and related problems, Theoret. Comput. Sci. 19(1982) 219-227.

[J] N. Jones and W. Laaser, Complete problems for deterministic polynomial time, Theoret. Comput. Sci. 3(1977) 105-117.

[K] J. Karhumäki, The Ehrenfeucht Conjecture: A compactness claim for finitely generated free monoids, Theoret. Comput. Sci. 29(1984) 285-308.

[KK] J. Karhumäki and H.C.M. Kleijn, On the equivalence problem of compositions of morphisms and inverse morphisms, RAIRO Inform. Théor 19(1985) 203-211.

[L] J. Lawrence, The non-existence of finite test sets for set-equivalence of finite substitutions, manuscript (1985).

[Ma] Y. Maon, Decision problems concerning equivalence of transductions on languages, Ph.D. Thesis, Tel Aviv University (1985).

[McN] R. McNaughton, A proof of the Ehrenfeucht Conjecture, Informal Memorandum (1985).

[Mo] E.F. Moore, Gedanken-experiments on sequential machines, in: Automata Studies (Princeton University Press, 1956).

[P] D. Perrin, On the solution of Ehrenfeucht's Conjecture, EATCS Bull. 27(1985) 68-70.

[RR] A. Restivo and C. Reutenauer, On cancellation properties of languages which are supports of rational formal power series, J. Comput. System Sci. 22(1984) 153-159.

[RS] G. Rozenberg and A. Salomaa, The Mathematical Theory of L Systems (Academic Press, New York, 1980).

[Ru] K. Ruohomen, Test sets for iterated morphisms, manuscript (1984).

[Sa] A. Salomaa, The Ehrenfeucht Conjecture: A proof for language theorists, EATCS Bull. 27(1985) 71-82.

[Sc] M.P. Schützenberger, Sur les relations rationnelles, in: Lecture Notes in Computer Science 33 (Springer, Berlin, 1975) 209-213.

[St] J.R. Stallings, Finiteness properties of matrix representations, manuscript (1985).