Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science
http://pic.sagepub.com/

Action aggregation and defuzzification in Mamdani-type fuzzy systems

D T Pham and M Castellani

Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 2002 216: 747
DOI: 10.1243/09544060260128797

Action aggregation and defuzzification in Mamdani-type fuzzy systems

D T Pham* and **M Castellani**
Manufacturing Engineering Centre, School of Engineering, University of Wales, Cardiff, Wales, UK

Abstract: This paper discusses the issues of action aggregation and defuzzification in Mamdani-type fuzzy systems. The paper highlights the shortcomings of defuzzification techniques associated with the customary interpretation of the sentence connective '*and*' by means of the set union operation. These include loss of smoothness of the output characteristic and inaccurate mapping of the fuzzy response. The most appropriate procedure for aggregating the outputs of different fuzzy rules and converting them into crisp signals is then suggested. The advantages in terms of increased transparency and mapping accuracy of the fuzzy response are demonstrated.

Keywords: fuzzy logic, output aggregation, defuzzification, COG method, MOM method, weighted average

NOTATION

a_Y	semi-support of membership function Y
$a_1(\alpha)$, $a_2(\alpha)$	extremes of the α-cut
A, B	fuzzy terms
A_n	area of nth fuzzy term of a fuzzy space partition
A_Y	area of fuzzy term Y
COG_n	centre of gravity of nth fuzzy term of a fuzzy space partition
$f(\mathbf{x})$	function
h	activation degree
h_Y	activation degree of fuzzy action Y
$T(\mathbf{x})$	transformation function
$\mathbf{x}$	variable
X_n, Y_n	fuzzy terms
$2a_Y$	support of triangular fuzzy term Y
α, β, γ, δ	parameters
$\mu_X(\mathbf{x})$, $\nu_X(\mathbf{x})$	membership value of element $\mathbf{x}$ of universe of discourse X

The MS was received on 18 December 2000 and was accepted after revision for publication on 21 March 2002.
* *Corresponding author: Manufacturing Engineering Centre, School of Engineering, University of Wales, Cardiff, PO Box 688, Newport Road, Cardiff CF24 0YH, Wales, UK.*

1 INTRODUCTION

In the design of fuzzy logic (FL) [**1**] systems, the Mamdani model [**2**, **3**] has its strong points in its closeness to Zadeh's method of fuzzy reasoning and for its human-like representation of the response policy. Being close to Zadeh's definition of FL, it allows a natural extension to the fuzzy domain of the familiar crisp modus ponens logical inferencing rule. As opposed to Takagi and Sugeno's model [**3**, **4**], Mamdani's model expresses the output using fuzzy terms instead of mathematical combinations of the input variables. The Mamdani model has its main shortcoming in its unsuitability for the analysis of closed-loop system stability. However, in many applications, this issue is not critical. For the above reasons, the Mamdani model is still the basis for many industrial FL applications and a full understanding of its properties would be of value in realizing further successful implementations and developments.

The focus of the paper is on the interactions between the output space partition, the rule aggregation operator and the defuzzification procedure. The basic properties of the aggregation and the defuzzification operations are summarized and a survey of output defuzzification procedures is presented. The limitations and advantages of the current procedures are discussed with particular regard to their transparency and mapping accuracy. Following the discussion, the most appropriate procedure for converting the output of the fuzzy rules into crisp signals is described. The suggested procedure allows full flexibility in the definition of the output partition, maintaining the transparency and predictability of the behaviour of the system. Moreover, the aggregation and defuzzification algorithm adopted in this work has a modular structure that lends itself to direct expression as a fuzzy neural network (FNN) architecture.

2 AN ANALYSIS OF MAMDANI-TYPE FL SYSTEMS

A block diagram of the information flow in a general Mamdani FL system is given in Fig. 1. The lower section of the 'FL System' block represents the fuzzy knowledge, while the upper part contains the knowledge processing operators.

The fuzzy knowledge base (KB) can be acquired from human expertise or automatically generated via machine learning techniques [**5**]. This static information is expressed in terms of fuzzy production rules mapping a set of conditions on to a set of actions, where each condition (action) is defined by the linguistic value of an input (output) variable. Individual rules are joined together by the sentence connective '*else*' to form the overall rule base (RB) [**2**]. Because rules involving multiple outputs can always be decomposed into a set of single-output rules [**5**], in the rest of this paper only rules defined over a one-dimensional output space will be considered. Accordingly, each rule consequent is meant to define one single-output action.

The partition of the input (output) space into a set of linguistic terms generates a symbolic description of the physical space. The 'link' between the real world and its linguistic expression is given by fuzzy membership functions (MFs) [**1**], which characterize the interpretation of the fuzzy input–output relationship.

The operators represented in the upper part of the 'FL System' block perform the symbolic processing of dynamic knowledge and the direct and inverse transformations between linguistic terms and numerical data. A desirable property of the fuzzy operators is a consistent behaviour irrespective of the nature of the stored knowledge. The absence of interactions between the processing algorithms and the KB is particularly important for the transparency and predictability of fuzzy systems.

The symbolic processing of fuzzy information is performed using the compositional rule of inference, which allows the determination of the activation of each rule consequent according to the degree of matching of the antecedent. Zadeh's *sup-min* [**6**] and Larsen's *sup-prod* [**5**] are the most popular composition operators. The overall fuzzy output is normally created by superposition of the individual rule actions. This implies the interpretation of the rule connective '*else*' with the union operation, which is normally implemented through the pointwise *max* operator [**5**]. Consequently, only information having the maximum activation value is used for generating the crisp signal, sometimes discarding the contribution of entire rules.

Two operators are needed, one to convert the input numerical data into qualitative information and the other to perform the inverse process on the qualitative output of the system [**6**]. In Fig. 1, these two operators have been represented as two logical blocks, the fuzzifier and the defuzzifier, added respectively at the input and the output of the fuzzy system. In the fuzzifier block, each numerical observation is mapped on to a fuzzy set (*direct mapping*). Most commonly, this fuzzy set is a fuzzy singleton. The fuzzified input is then matched with the rule antecedents and a set of fuzzy actions is generated. The overall fuzzy action is expressed as a possibility (truth) distribution over the universe of

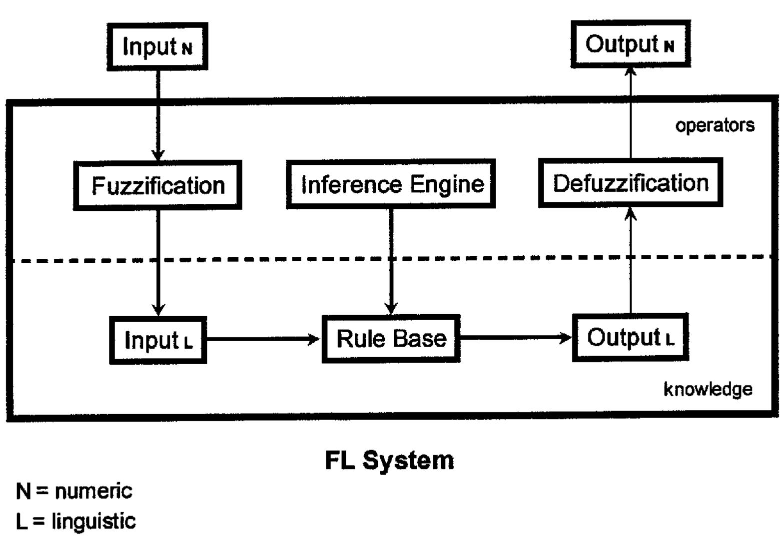

Fig. 1 Mamdani FL System: information flow

discourse and it needs to be converted into some numerical output.

There is not yet a commonly accepted procedure for converting qualitative information into a numerical form (*inverse mapping*). Unlike the fuzzification process, which is a one-to-one mapping from the input space to the possibility interval, the defuzzification procedure poses the problem of finding a suitable many-to-one mapping from a possibility distribution to the output space. The choice of the crisp value best representing such a possibility distribution is normally affected by factors such as problem requirements and implementation constraints. The two most commonly used defuzzification procedures in FL are the *centre of gravity* (COG) method and the *mean of maxima* (MOM) method [**5, 7**], both having advantages and shortcomings that will be discussed later. Alternative methods include Yager and Filev's *basic defuzzification distribution* (BADD) [**8, 9**], *semi-linear defuzzification* (SLIDE) and *modified-SLIDE* (M-SLIDE) [**10**], Jiang and Li's *Gaussian distribution transformation-based defuzzification* (GTD), *polynomial transformation-based defuzzification* (PTD) [**11**] and *multimode-oriented PTD* (M-PTD) [**12, 13**], Runkler and Glesner's *extended centre of area* (XCOA) [**14**] and Saade's approach unifying defuzzification with the comparison of fuzzy sets [**15**]. Mizumoto [**16**] compares the influence of several rule aggregation and output defuzzification methods on the control response of a simulated plant, and presents two algorithms, the *height method* and the *area method*, that bring together aggregation and defuzzification of fuzzy actions. The height method has also been adopted by Cherkassky and Mulier [**17**] under the name *additive defuzzification*.

The overall crisp relationship mapped by the 'FL System' block of Fig. 1 is therefore not only dependent on the KB definition but also on the mathematical operators chosen to manipulate it [**18**].

3 AGGREGATION AND DEFUZZIFICATION OF RULE ACTIONS

3.1 Aggregation of fuzzy actions

In a fuzzy system, once the conditions of the rules have been matched, a set of actions is activated. Each rule whose antecedent has a non-zero matching degree will contribute an output with an activation value equal to the matching degree of the antecedent. The *max* operator is by far the most common implementation of the rule aggregation operation. According to this procedure, the overall fuzzy output is calculated from the set of individual outputs taking the maximum truth value where one or more terms overlap. Figure 2 shows an example of *max* aggregation for two overlapping actions A and B with activation degrees of 0.3 and 0.8 respectively.

Alternative aggregation operations have been proposed generally based on a different implementation of the union operation. The most common examples replace the *max* operator with other triangular co-norms such as the *algebraic sum* or the *bounded product* [**5, 16**]. As the ordering of the rules is unimportant, any operation possessing the properties of commutativity and associativity is a candidate for implementing the *else* connective. In practice, the quality of the choice is

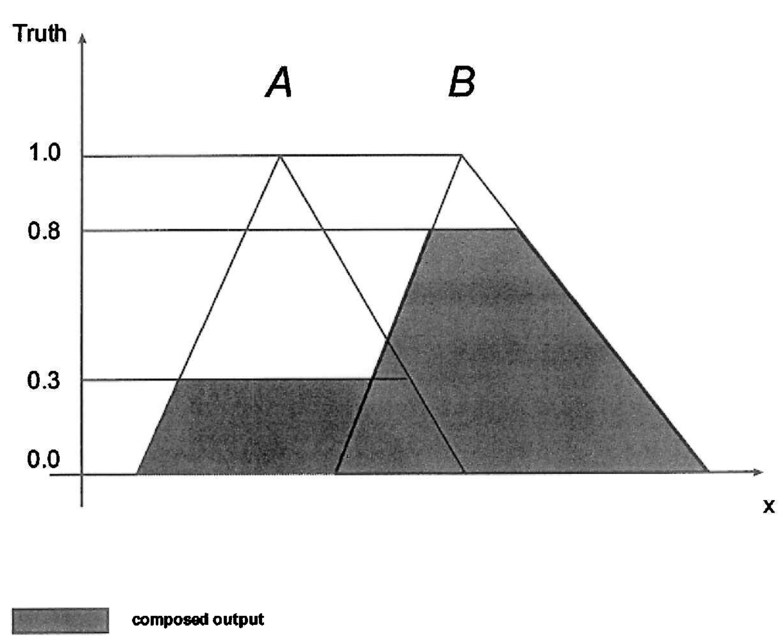

Fig. 2 *sup* composition

normally affected by the type of composition rule used. Even though several studies support the association of the union operation to the *sup-star* composition rule (see reference [**5**] for an overview), successful results have been reported using an additive procedure [**16, 17**].

3.2 Output defuzzification

The overall fuzzy output generally constitutes a multimodal non-zero truth distribution of possible crisp values over a subset of the output space. In the defuzzification stage, one of those possible crisp values has to be selected as the output crisp signal. The design of a sound defuzzification method is important as it will affect the interpretation of the fuzzy response policy. The first aim is always to be able to generate a crisp value that will be representative of the output possibility distribution.

For this purpose, some desirable defuzzification properties are outlined in reference [**7**], of which the basic ones are consistency, section invariance and monotonicity. A consistent defuzzification method maps convex crisp sets to their centroid. From this, it follows that the empty set and the universal set are both defuzzified to their centre and a fuzzy singleton is defuzzified to its sole non-zero truth element. Section invariance guarantees that the defuzzified value is uniquely dependent on the output space elements having a non-zero truth value. Outside this set, any modification of the fuzzy universe of discourse does not affect the defuzzified value. Monotonicity requires that, for any decrease (increase) in the truth degree of a single-output space element, the defuzzification result remains unchanged or is moved away from (closer to) that element. This property implies that the contribution of each single-output space element to the final defuzzified value increases with the degree of truth of that element.

A desirable defuzzification procedure should also require a low computational effort to allow its implementation in real-time applications. At the same time, it should allow a smooth response and mapping accuracy to be obtained over all or most of the output space.

Finally, an ideal defuzzification method should ease the design of the fuzzy system and keep the decision-making logic transparent to the user. The manipulation operators should not add any overhead to the system design and analysis.

Any defuzzification method possessing the abovementioned properties is a good candidate for implementation in an FL expert system. The following section provides a critical overview of some of the most common defuzzification methods. For an application-oriented overview of defuzzification methods the reader is referred to reference [**19**].

4 MAIN DEFUZZIFICATION TECHNIQUES

4.1 Maxima methods

Once the shape of the output possibility distribution has been determined, a quick and simple defuzzification procedure is to pick up one of the crisp values having a maximum truth degree (*maxima* methods) [**7, 20**]. Possible choices are the first (smallest), the last (largest) or, in the case of a unimodal possibility distribution, the median value. By far the most common maxima method is to select the mean value of elements with maximum truth degrees (MOM method) [**5, 7**].

The strength of maxima methods lies in their simplicity and speed of execution, but their major weakness is in not being truly fuzzy. They are section invariant and monotonous, but as a consequence of only considering elements of highest membership degrees, information not related to rules of maximal activation is ignored. This causes a loss of the smooth output characteristic generated by the gradual transitions from input space areas where different rules prevail [**18**]. The crisp response curve does not retain the continuous and gradual nature of the input–output fuzzy relationship and it is characterized by brisk discontinuities of the kind produced by a multilevel relay system. The functional identity between a multilevel relay and an FL system using symmetrical MFs, the *sup-star* composition rule and the MOM defuzzification procedure has been demonstrated in reference [**21**].

For control applications, the type of multirelay control characteristic induced by maxima procedures may also show some of the limitations of such systems. Suboptimal control performances in terms of steady state error, minimization of control effort and plant fluctuations have occasionally been associated with the use of the MOM method [**22, 23**]. To retain the smooth output transitions typical of FL control, it is necessary for the defuzzification procedure to consider also values whose membership degree is other than maximal. The most popular alternative to maxima methods is to calculate the final crisp value from the area of the output possibility distribution (*area-based* methods).

4.2 Area-based methods

A well-known area-based procedure is to select the defuzzification result as the centroid of the output possibility distribution COG method [**5, 7**]. A close variant is the *centre of area* (COA) method [**7**], which selects the crisp value as the position where the output area can be split into two equal halves. If the output distribution is symmetrical, the two methods give identical results.

The COG defuzzification method is section invariant, monotonous and consistent, and its deterministic response curve is characterized by a smooth and continuous behaviour. In control applications, compared to maxima methods, the COG rule gives a superior steady state performance and a reduction of control effort and plant oscillations [**22, 23**]. On the other hand, a better transient performance using the MOM defuzzification procedure has been reported [**22**]. This is probably because the output of the MOM method is characterized by fewer but larger variations, corresponding to the boundaries between neighbouring rules. Given the same RB, this allows larger corrections, which should explain the superior transient response.

4.3 Other methods

Modelled on the application of the COG procedure, several learning algorithms have been proposed for output defuzzification. The general idea underlying all these methods is to perform some transformation of the output possibility distribution according to an automatically generated set of parameters. A generic transformation on a fuzzy term X can be written as [**11**]

$$v_X(\mathbf{x}) = \mu_X(\mathbf{x}) T(\mathbf{x}) \tag{1}$$

where $v_X(\mathbf{x})$ is the new membership value of the element $\mathbf{x} \in X$, $\mu_X(\mathbf{x})$ is the original value and $T(\mathbf{x})$ is any transformation function. Some examples of such functions are:

BADD method [**8, 9**]:

$$T(\mathbf{x}) = \mu_X(\mathbf{x})^{\gamma - 1} \tag{2}$$

where γ is an automatically learned parameter.

SLIDE method [**10**]:

$$T(\mathbf{x}) = \begin{cases} 1 - \beta, & (\mathbf{x} \leqslant \alpha) \\ 1, & (\mathbf{x} > \alpha) \end{cases} \tag{3}$$

where α and β are parameters to be learned.

M-PTD method [**11, 12**]:

$$T(\mathbf{x}) = \left[\sum_{j=0}^{N} \beta_j (\mu_X(\mathbf{x}) - 0.5)^j \right]^2 \tag{4}$$

where β_j are parameters that are adaptively tuned and $\mu_X(\mathbf{x})$ and $T(\mathbf{x})$ are discrete functions.

In all of the above procedures, the kind of transformation induced by $T(\mathbf{x})$ is a distortion of the output possibility distribution in order to increase or decrease the weight of elements of higher membership values. The more the transformation magnifies the differences between low and high membership values, the more the defuzzification method will approximate the MOM characteristic. In the limiting case, all values except the maximal will be brought to zero and the defuzzification procedure will correspond to the MOM method. On the other hand, if the transformation is the identity function, the defuzzification procedure will correspond to the COG method. A similar approach has been adopted in reference [**15**], where the output possibility distribution is defuzzified to the value

$$\mathbf{y} = \int_0^1 [\delta a_1(\alpha) + (1 - \delta) a_2(\alpha)] \, \mathrm{d}\alpha \tag{5}$$

In equation (5), δ is a parameter set by the designer and $a_1(\alpha)$ and $a_2(\alpha)$ are the two extremes of an α-cut $X_\alpha = \{a \in U | \mu_X(a) \geqslant \alpha\}$ [**3**] of output possibility distribution X defined over universe of discourse U. The aim of this procedure is to magnify the contribution of the elements on one side of the centroid of the output distribution. In this way, more or less drastic actions can be achieved.

These algorithms can help to tune the output of the system but they also introduce a considerable computational overhead for the determination of the transformation parameters and in some cases in the defuzzification procedure. The action of some of the parameters is also not immediately obvious and this can affect the transparency of the fuzzy system. It is preferable whenever possible to tune the behaviour of the system directly by acting on the shape of the output MFs. The response curve of the system can be made more or less smooth by adjusting the overlap of the output terms, and efforts towards an adaptive behaviour should be focused in the same direction. Changes in the behaviour of the system should reflect variations in the response policy that should be encoded in the RB and the input and output space partitions. Keeping the fuzzy knowledge conceptually separate from the operators can considerably improve the transparency and predictability of the fuzzy system.

5 COMPARISON OF DEFUZZIFICATION RESULTS

To illustrate the influence of different defuzzification procedures on the overall system response, a general single-input–single-output Mamdani-type fuzzy system is used for modelling a linear crisp relationship. The task has been chosen for the straightforwardness of the KB design and the ease of detecting any divergence from the desired behaviour. As multidimensional fuzzy spaces are generated by performing the Cartesian product of their one-dimensional components [**5**], the results achieved in the example can be readily extended to more complex multi-input fuzzy rules.

The system adopts Zadeh's *sup-min* composition rule of inference and interprets the logical connectives *else* and *and* respectively using the *max* and *min* operators. By varying the defuzzification method and keeping the

Table 1 Rule base

Rules	In	Out
1	X_0	Y_0
2	X_1	Y_1
3	X_2	Y_2
4	X_3	Y_3
5	X_4	Y_4

rest of the system unaltered, it is possible to compare the effect of different defuzzification procedures.

The approximate fuzzy mapping is defined by partitioning the input and output spaces each into five fuzzy terms, respectively X_0–X_4 and Y_0–Y_4, delimited by triangular MFs equally spaced between consecutive peaks. Differently from many FL implementations, in this experiment the two terms at the extremes of the output universe of discourse keep a symmetrical triangular shape and their peaks do not correspond to the space boundaries. This configuration avoids certain deformations of the input–output characteristic that will be discussed later. The partition of the input and output spaces is *normalized* [**3**]; i.e. at any point of the universe of discourse no more than two fuzzy sets overlap and the sum of their membership degrees is always equal to 1. The fuzzy mapping is realized by defining the five rules listed in Table 1.

The overall fuzzy associative memory (FAM) [**24**] and the desired linear crisp behaviour are illustrated in Fig. 3. It is possible to see that the definition of the fuzzy map represents a direct fuzzification of the crisp linear relationship. Ideally, the fuzzy system should therefore be able to reproduce the desired response curve.

Figure 4 displays the crisp input–output relationships produced by two fuzzy systems using respectively the MOM and COG rules. The input and output fuzzy partitions are shown next to the input–output coordinate axes and the desired linear characteristic is plotted for reference. The figure illustrates the stepwise output response obtained using the MOM method and already documented in reference [**21**] together with the smoother response given by the COG rule.

The application of the COG method allows a better reproduction of the desired output, but introduces an oscillatory behaviour around the reference curve. The reason for this inaccuracy is that the contribution of each term to the final crisp value is determined by the portion of its area aggregated into the overall fuzzy output (see Fig. 2). Fuzzy output terms having a high activation level will contribute with larger sections of their total area, therefore driving the final crisp output closer to their centre. Unfortunately, the section of activated area does not increase linearly with the firing strength of an output term, but in convex MFs it grows more quickly for low firing strengths. In the case of triangular MFs, as used in the example of Fig. 4, the activated area contributing to the total fuzzy output is, for each term Y_i, equivalent to

$$A_{Y_i} = (2 - h_{Y_i})a_{Y_i}h_{Y_i} = -a_{Y_i}h_{Y_i}^2 + 2a_{Y_i}h_{Y_i} \qquad (6)$$

where A_{Y_i} is the activated area, h_{Y_i} is the activation level of Y_i and a_{Y_i} is half the support of the MF. According to equation (6), the weighting factor (i.e. the portion of area) for the contribution of each term to the defuzzified output grows parabolically with the activation degree.

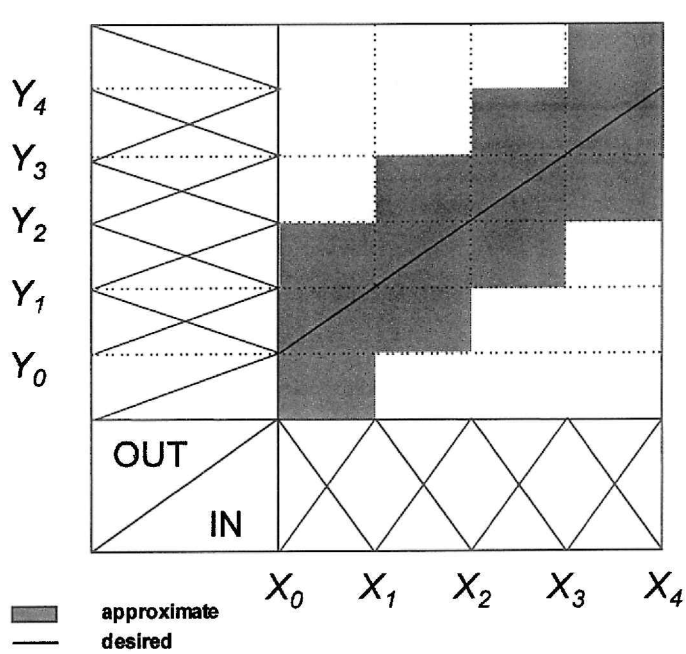

Fig. 3 Fuzzy map and desired behaviour

The contribution of each rule consequent will therefore rise quickly and fade slowly as the input value crosses the 'attraction basin' of the rule. This determines the distortion of the response of the system towards the average values between the centroids of the activated terms. In the example of Fig. 4, the response is biased towards the intermediate values between consecutive centroids, alternately overshooting or undershooting the reference output.

A more severe drawback follows from scaling the contribution of each action according to the area of its possibility distribution. Considering, for example, two non-overlapping actions triggered by a certain set of conditions, the defuzzified value will be the centroid of the system composed of their two possibility distributions. From a well-known property of the COG, this is equivalent to the centroid of a system composed of two elements, each placed on the centroid of one of the two distributions and having a possibility equal to the area of that distribution:

$$\mathrm{COG} = \frac{\mathrm{COG}_1 A_1 + \mathrm{COG}_2 A_2}{A_1 + A_2} \tag{7}$$

where COG, COG_1, A_1, COG_2 and A_2 are respectively the overall centroid position, the centroid position and area of the first distribution, and the centroid position and area of the second distribution. From the above formula, it is clear that the defuzzified value will be closer to the action whose possibility distribution has the largest area. The rationale for this is the attempt to bring the output closer to the action of the rule of maximal activation; this will happen as long as the output space has been partitioned by a set of MFs of equal areas. In this case, the action whose possibility distribution has the largest area will correspond to the action belonging to the rule of maximal activation.

However, if different output linguistic terms have possibility distributions of different areas, the COG method will no longer guarantee a defuzzified value close to the action of maximal activation. Actions whose MF encompasses a large area will dominate the response curve in a way that is proportional to the size of their area. A consequence is that output terms having MFs defined with a high vagueness (fuzziness) will contribute with a large area and consequently dominate the output of the fuzzy system. Considering, for example, an action Y defined by a triangular MF of base $2a_Y$ (i.e. the support) and firing strength h_Y, its contribution A_Y to the final crisp output is determined by equation (6). A_Y increases with h_Y in a parabolic way, but it also grows proportionally to a_Y. In a triangular MF, a_Y gives a measure of the fuzziness of the linguistic term—the fuzzier the definition, the larger a_Y. For instance, if two actions have the same firing strength and A_1 is four times larger than A_2, the defuzzified output will be four times closer to COG_1 than to COG_2. If A_1 is four times larger than A_2 and the firing strength h_2 of the second rule is equal to 1, then according to equation (6) a firing strength h_1 equal to 0.13 would be enough for the first

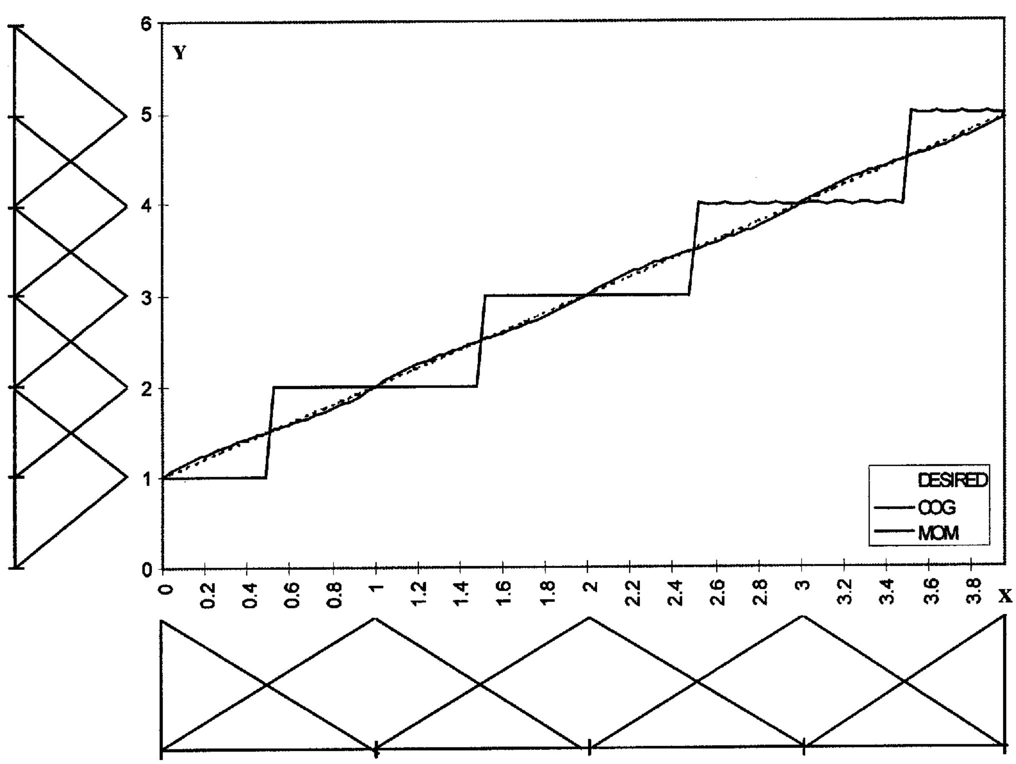

Fig. 4 MOM method and COG rule

rule to bring the defuzzified output to the mid-point between COG_1 and COG_2.

In standard Mamdani-type FL systems where the *max* operator is customarily used for action aggregation, the application of the COG method has therefore the undesirable effect of weighting the linguistic outputs proportionally to their imprecision. Moreover, the smoothness of the output transition between two neighbouring rules is affected by the MF area mismatch between their rule actions. Rule actions having possibility distributions of large area will prevail until their activation level becomes very low, at which point the output of the system briskly changes to the value indicated by the action of the neighbouring rule.

Figure 5 illustrates this for an input–output relationship similar to the one plotted in Fig. 4. The numbers of input and output MFs have been decreased to three, the reference output has been kept as a straight line and the RB has been reduced to the first three rules of Table 1. Compared to the MF of the other two actions, the MF of Y_1 in the second rule encompasses an area four times larger. The output of the MOM method has been kept to show the boundaries between the areas where different rules prevail. The plot reveals the dominance of the second output term well beyond the points where the 'steps' in the MOM response mark the end of the area of prevalence of the second rule.

Saade [15] has pointed out the fact that the totality of the crisp output range is not used as a weakness of the COG method. The closest the output can approach the boundaries of its interval of definition are in fact the centroids of the two MFs at the extremes of the fuzzy partition. Unless fuzzy singletons are employed, these values will never coincide with the output range extremes. Therefore, there are always two bands of values at the bottom and at the top of the output range that are not 'reachable' by the fuzzy system (see also reference [7]). Figure 6 shows this situation for a fuzzy system partitioning the input and the output spaces into three linguistic terms each and using the first three rules of Table 1 to model the usual linear relationship. To maximize the response interval, the peak values of the two MFs at the extremes of the action range have been set to the output interval extremes. The area under the second output MF is twice the area of the other two terms in order to maintain a normalized partition of the output universe of discourse.

In an approximate mapping, the feature of rarely recommending actions close to the extremes of the output range can generally be considered desirable, as it avoids extreme responses in the presence of uncertainty. However, problems arise when attempting to compensate a possible slower transient response by adjustments of the fuzzy mapping. To enlarge the action range, it is necessary to bring the centroid of the extremal MFs

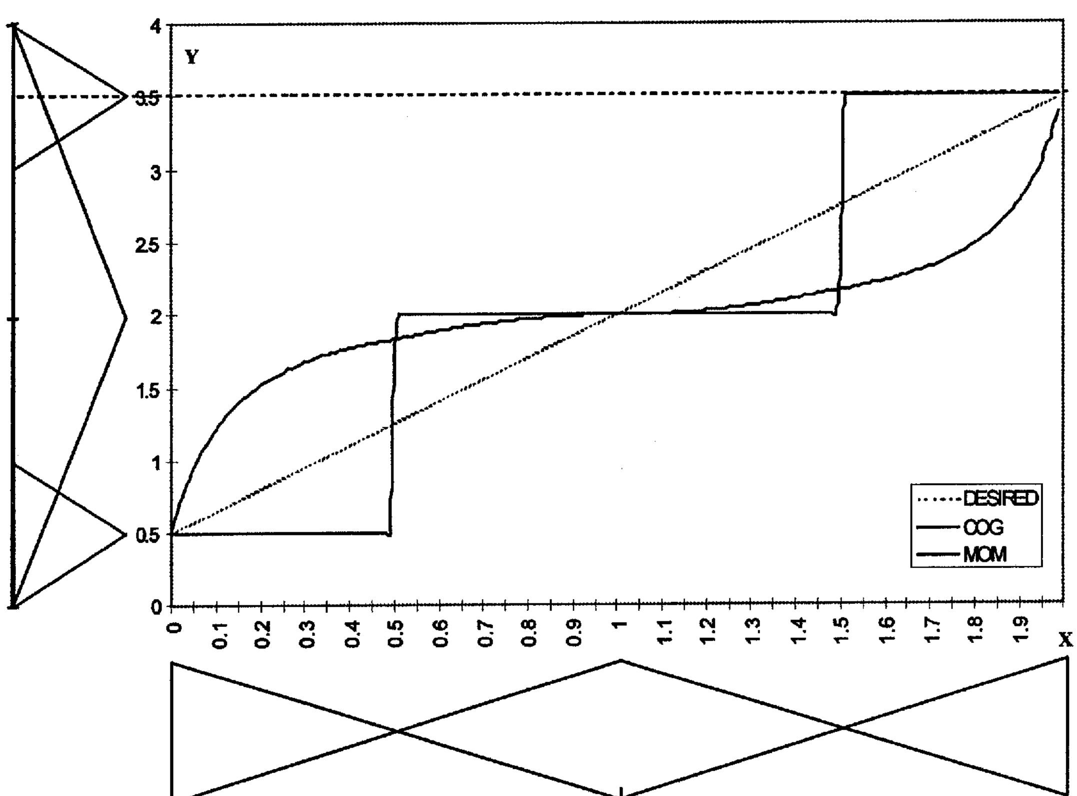

Fig. 5 Output curve distortion: MF of Y_1 is 4 times wider than those of Y_0 and Y_2

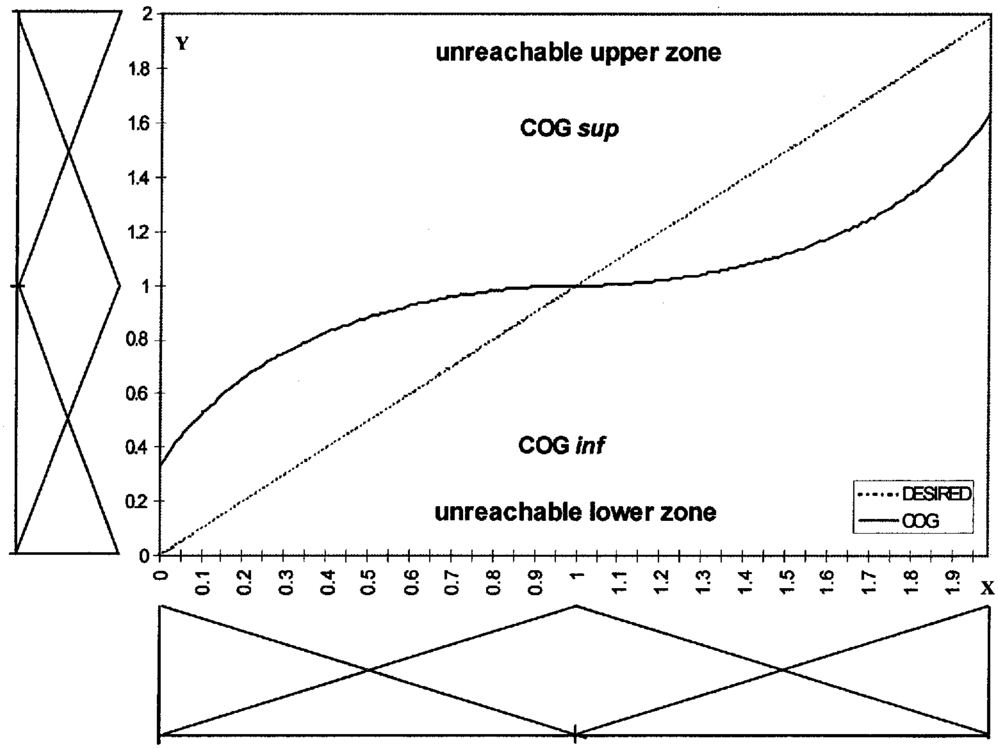

Fig. 6 COG defuzzification range

closer to the limits of the output range. This can be accomplished by reducing the width and hence the area of the extreme MFs. Unfortunately, as shown above, the response characteristic is affected by the relative areas encompassed by the output MFs. Reducing the areas of the two extreme terms will therefore cause a distortion of the response curve.

Figure 6 shows the kind of deformation introduced when the support of the two extreme MFs is half the width of the other term. To obtain a smoother output characteristic, a finer space partition has now to be introduced to reduce the mismatch between the area of the extreme output terms and the area of the third term. The designer will therefore have to enlarge the size of the RB, increasing the complexity of the system and affecting therefore its transparency and speed of execution.

The use of fuzzy singletons to enlarge the output range would limit the flexibility of the system design. In particular, it would be impossible to modulate the response of a single rule according to its activation degree by conveniently shaping the MFs of the rule actions. Moreover, there is a sizeable amount of literature featuring non-singleton fuzzy MFs where the introduction of such a constraint would imply a major review of the interpretation of the fuzzy algorithm.

The BADD, SLIDE and M-PTD defuzzification methods will approach the behaviour of either the MOM or COG method according to the tuning of the transformation parameters. Saade's method is essentially an area-based procedure and as such it will behave similarly to the COG method. None of the procedures listed in Section 4.3 will therefore overcome the shortcomings affecting the MOM and COG methods.

6 AN IMPROVED INTERPRETATION OF FUZZY RULE ACTIONS

The above analysis showed the inadequacy of maxima procedures and the drawbacks of area-based methods when combined with the implementation of action aggregation through the *max* operator. For the process of output defuzzification, area-based methods seem to be the most appropriate choice due to their properties of consistency, section invariance and monotonicity and the straightforwardness of the algorithm. Nonetheless, it is necessary to find an alternative aggregation procedure to avoid the weighting of the contribution of each output according to the area of its possibility distribution.

It is clear that changing the implementation of the union operation does not serve the purpose, as it would only modify the way of forming the pointwise combinations of the possibility values of the actions in the overall output distribution. It is therefore necessary to change the process of MF superposition related to the operation of set union. To this end, Mizumoto [**16**] and Cherkassky and Mulier [**17**] proposed that the aggregation operator be implemented through an additive procedure.

The idea behind these methods is to compute the final crisp value from the properties of each separate action rather than from an aggregated fuzzy output [**16**]. Each action term is first defuzzified into a representative point and then combined with the values of the other actions to work out the final crisp output. The COG defuzzification method is still used for the choice of each action representative and for the combination of the action centroids in the overall crisp output. Each action representative is weighted by a factor related to the firing strength of the rule. The overall process can therefore be divided into two cascaded operations of centroid defuzzification. The first defuzzification is applied to obtain the discrete fuzzy set composed of the action centroids and their activation degrees, while the second defuzzification operation yields the final output out of that set. The defuzzified output COG can be written as

$$\mathrm{COG} = \frac{\sum_i \mathrm{COG}_i f(h_i)}{\sum_i f(h_i)} \tag{8}$$

where COG_i and $f(h_i)$ are respectively the centroid position and a general function of the activation degree of the ith term. If the function f is the activation-dependent area of the output possibility distribution, the defuzzification becomes Mizumoto's area method. If f corresponds to the identity function, the operation is referred to as the height method [**16**] or additive defuzzification [**17**].

Because the scaling factor of each rule action is still the area of its possibility distribution, the area method cannot be expected to overcome the drawbacks of other area-based methods. The height method is instead a good algorithm for the process of action aggregation and defuzzification.

In the computation of the crisp output, the height method does not take into account the areas of the MFs of the actions. Only the rule activation degree appears in the calculation of the contribution of each action. Therefore, terms of wide possibility distribution will not dominate the output of the controller and will not affect the smoothness of the response curve. At the same time, at the level of each individual rule, the centroid-based procedure allows the tuning of the location of the representative point of the action according to the degree of activation of the rule. Defining the output terms through asymmetric possibility distributions allows the shifting of the COG of the rule output according to the activation degree.

Figure 7 shows the response curves of two fuzzy systems using respectively the COG and the height methods. The same single-input–single-output example of Fig. 4 was employed to enable the results to be plotted. However, the procedure adopted is general and valid for any n-dimensionally complex input space. It is possible to see that the output characteristic of the FL system using the height method coincides with the

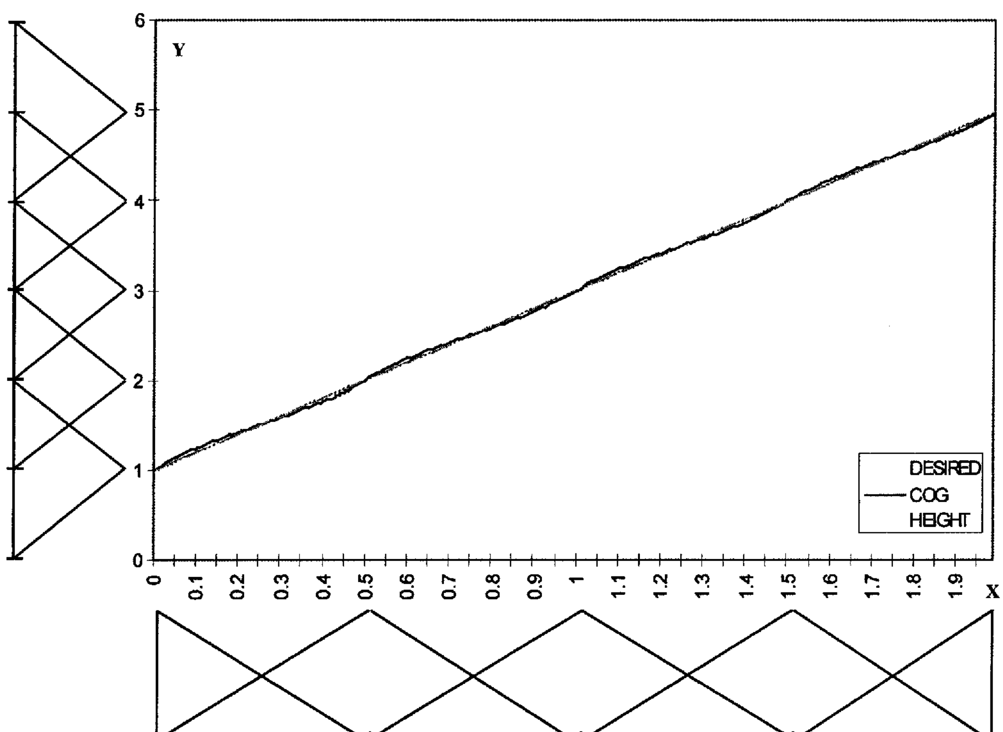

Fig. 7 COG rule and height method

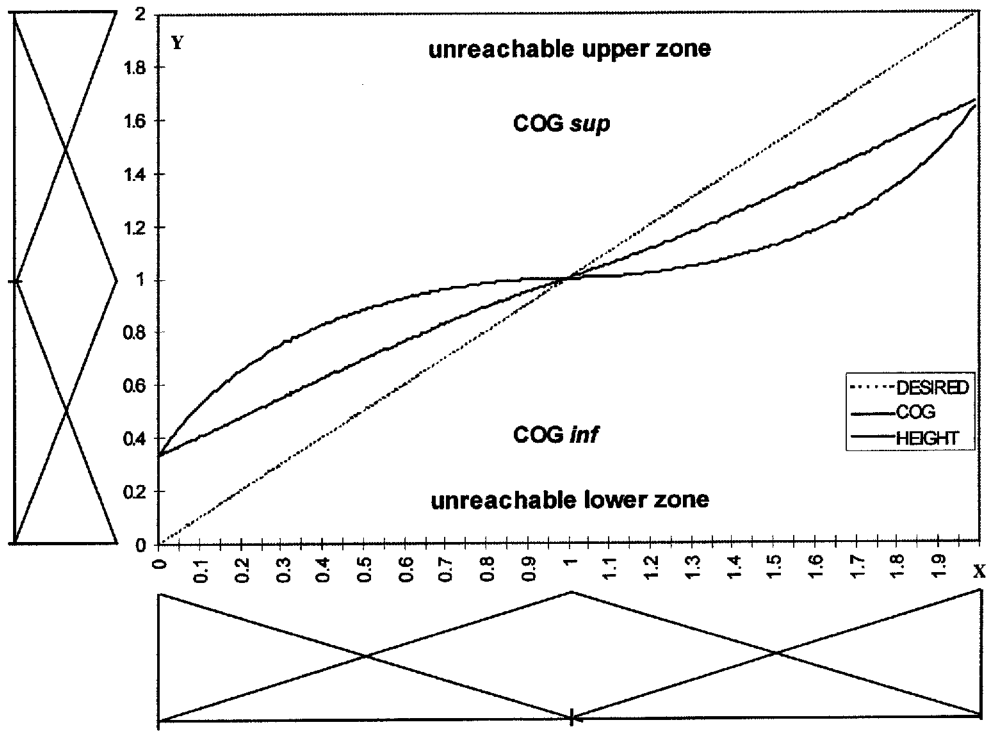

Fig. 8 COG rule and height method: defuzzification range

desired straight line. In particular, its response curve is not affected by the oscillatory behaviour caused by the combination of the COG method with the *max* aggregation operator.

Moreover, the system response is now not distorted towards the centroids of MFs with larger areas. Figure 8 shows the results of applying the height method and the standard union aggregation and COG defuzzification procedure to the same example as Fig. 6. Even though the MF of the second rule action encompasses an area twice as large as the other two output terms, the response curve obtained using the height method remains undistorted and follows a straight line.

The defuzzification range of the height method does not improve upon the one obtainable by combining union aggregation and COG defuzzification. The system's output still does not cover the full output scale and the extreme output values are not reached. Because of this limited output range, the slope of the straight line plotted in Fig. 8 is smaller than the desired one.

Nevertheless, it is now possible to widen the output range, while restricting the width of the extremal MFs without further distorting the output characteristic. Figure 9 shows the result of reducing the support of the extreme MFs by a factor of 10. The response obtained using the height method is compared with the output of a standard fuzzy system using *max* aggregation and COG defuzzification. It is possible to see how the former is now almost identical to the desired behaviour, while the latter has an almost flat characteristic with two sharp steps at the extremes of the sampling space.

The height method therefore gives substantial advantages in action aggregation and defuzzification in FL systems. It is interesting to note that none of the authors who investigated this method have noted such advantages. In reference [**16**] various aggregation and defuzzification methods were compared on the basis of the results obtained for the fuzzy control of a simulated plant. The study was limited to the performance of the fuzzy controller, without giving any analysis of the reasons underlying the results. Cherkassky and Mulier [**17**] proposed the height method for its ease of implementation and system analysis. This is particularly true when symmetrical output MFs are used. In this case, the representative point of each fuzzy action is constant and equal to the middle point of the support of the term. Each action term can therefore be described through a single parameter corresponding to the centre of its MF, allowing a considerable simplification of the architecture of the system. Under this form, the fuzzy inferencing becomes equivalent to a special case of the Takagi–Sugeno model, where the linear combinations of the input variables are replaced by constant values in the rule consequents. Because of its ease of implementation, this particular representation has often been adopted in FNN applications.

The main limitation of the height method in comparison with the union aggregation and COG defuzzification procedure is in the reduced use of the

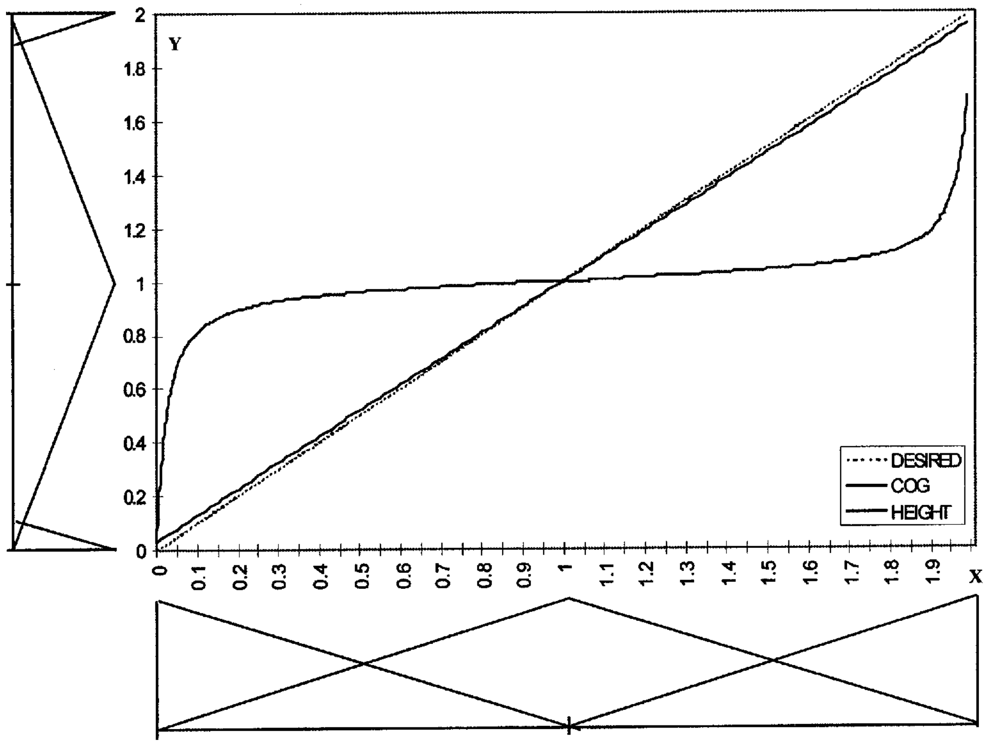

Fig. 9 Extended output range

information related to the MF shape. This information is in fact utilized now only to determine the representative point for each individual term. In particular, since the contribution of each action is no longer dependent on the area of its possibility distribution, differences between the MF shape of different actions cannot now be used to tune the fuzzy output.

7 CONCLUSIONS AND FURTHER WORK

This paper has focused on the operations of aggregation and defuzzification of fuzzy rule actions in Mamdani-type fuzzy systems. The superior transparency and mapping accuracy of the height method with respect to the usual combination of the union of fuzzy sets with maxima or area-based methods have been demonstrated. Because of the modular way in which the crisp output is formed, the height method is also well suited to an FNN implementation.

The above conclusions were drawn independently of the implementation of the fuzzy inferencing operators and the compositional rule of inference (e.g. *sup-max*, *sup-prod*, etc). The results presented are therefore valid for a wide range of problems. The only constraint in applying the proposed procedure is that the sentence connective *else* should be generally interpreted through the union operation.

Further work should be directed at studying the effects of the shape of the output MFs on the overall crisp output and its interaction with the action aggregation and defuzzification procedures. Additional studies should also be conducted on the issue of closed-loop stability when applying the proposed FL method with asymmetrical output MFs.

REFERENCES

1 **Zadeh, L. A.** Fuzzy sets. *Inf. Control*, 1965, **8**, 338–353.
2 **Mamdani, E. H.** Application of fuzzy algorithms for control of simple dynamic plant. *Proc. IEE*, 1974, **121**(12), 1585–1588.
3 **Fantuzzi, C.** Bases of fuzzy control. In Proceedings of International Summer School on *FLC Advances in Methodology*, Ferrara I, 1998, pp. 1–34.
4 **Takagi, T.** and **Sugeno, M.** Fuzzy identification of systems and its application to modelling and control. *IEEE Trans. on Syst. Man and Cybernetics*, 1985, **15**, 116–132.
5 **Lee, C. C.** Fuzzy logic in control systems: fuzzy logic controller. Part I and Part II. *IEEE Trans. Syst. Man and Cybernetics*, 1990, **20**(2), 404–418 and 419–435.
6 **Zadeh, L. A.** Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Syst. Man and Cybernetics*, 1973, **3**(1) 28–44.
7 **Runkler, T. A.** Selection of appropriate defuzzification methods using application specific properties. *IEEE Trans. on Fuzzy Syst.*, 1997, **5**(1), 72–79.

8 **Yager, R. R.** and **Filev, D. P.** A generalized defuzzification method under BADD distributions. *Int. J. Intell. Syst.*, 1991, **6**, 689–697.

9 **Yager, R. R.** and **Filev, D. P.** On the issue of defuzzification and selection based on a fuzzy set. *Fuzzy Sets and Syst.*, 1993, **55**, 255–271.

10 **Yager, R. R.** and **Filev, D. P.** SLIDE: a simple adaptive defuzzification method. *IEEE Trans. on Fuzzy Syst.*, 1993, **1**, 69–78.

11 **Jiang, T.** and **Li, Y.** Generalized defuzzification strategies and their parameter learning procedures. *IEEE Trans. on Fuzzy Syst.*, 1996, **4**(1), 64–71.

12 **Jiang, T.** and **Li, Y.** Multimode-oriented polynomial transformation-based defuzzification strategy and parameter learning procedure. *IEEE Trans. on Syst. Man and Cybernetics*, 1997, **27**(5), 877–883.

13 **Jiang, T.** and **Li, Y.** Techniques and applications of fuzzy theory in generalised defuzzification methods and their utilisation in parameter learning techniques. In *Fuzzy Theory Systems, Techniques and Applications*, 1999, Vol. 2, pp. 872–896 (Academic Press, New York).

14 **Runkler, T. A.** and **Glesner, M.** Defuzzification with improved static and dynamic behaviour: extended center of area. In Proceedings of EUFIT, First European Congress on *Fuzzy and Intelligent Technology*, Aachen, Germany, September 1993, pp. 845–851.

15 **Saade, J. J.** A unifying approach to defuzzification and comparison of the outputs of fuzzy controllers. *IEEE Trans. on Fuzzy Syst.*, 1996, **4**(3), 227–237.

16 **Mizumoto, M.** Improvement methods of fuzzy control. In Proceedings of 3rd International Conference on *IFSA*, August 1989, pp. 60–62.

17 **Cherkassky, V.** and **Mulier, F.** *Learning from Data, Concepts, Theory and Methods*, 1998 (John Wiley, New York).

18 **Braae, M.** and **Rutherford, D. A.** Fuzzy relations in a control setting. *Kybernetes*, 1978, **7**, 185–188.

19 **Cox, E.** *The Fuzzy Systems Handbook*, 2nd edition, 1999 (Academic Press, New York).

20 **Zadeh, L. A.** Fuzzy algorithms. *Inf. Control*, 1968 **12**, 94–102.

21 **Kickert, W. J. M.** and **Mamdani, E. H.** Analysis of a fuzzy logic controller. *Fuzzy Sets and Syst.*, 1978, **1**, 29–44.

22 **Scharf, E. M.** and **Mandic, N. J.** The application of a fuzzy controller to the control of a multi-degree-of-freedom robot arm. In *Industrial Applications of Fuzzy Control* (Ed. M. Sugeno), 1985, pp. 41–61 (Elsevier–North-Holland).

23 **Larkin, L.** A fuzzy logic controller for aircraft flight control. In *Industrial Applications of Fuzzy Control* (Ed. M. Sugeno), 1985, pp. 41–61 (Elsevier–North-Holland).

24 **Kosko, B.** *Neural Networks and Fuzzy Systems*, 1992 (Prentice-Hall, Englewood Cliffs, New Jersey).