Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Observations on auditory learning in amplitude- and frequency-modulation rate discrimination

Pablo F. Hoffmann

Acoustics, Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg Ø, Denmark

PACS: 43.66.Fe

ABSTRACT

Because amplitude- and frequency-modulated sounds can be the basis for the synthesis of many complex sounds, they can be good candidates in the design of training systems aiming at improving the acquisition of perceptual skills that can benefit from information provided via the auditory channel. One of the key issues when designing such training systems is in the assessment of generalization of learning. In this study we present data on the learning of an auditory task involving sinusoidal amplitude- and frequency-modulation tones. Modulation rate discrimination thresholds were measured during pre-test, training, and post-test phases. During training, listeners were divided into two groups; one group trained on amplitude-modulation rate discrimination and the other group trained on frequency-modulation rate discrimination. Results showed certain degree of specificity for the trained conditions, differences in learning rate, and generalization across modulation type.

INTRODUCTION

Auditory learning is part of the more general concept of perceptual learning that refers to the improvement in performance as a result of training or practice in perceptual tasks (Lu et al. 2009). Several behavioral studies have shown that training improves performance on a multitude of auditory tasks (Demany 1985; Grimault et al. 2003; Karmarkar and Buonomano 2003; Amitay, Hawkey, and Moore 2005; Fitzgerald and Wright 2005; Zhang, Travis, and Collings 2007; Zhang and Wright 2009; Kumpik et al. 2009). A significant question about learning is whether the benefits of the acquired knowledge are specific to the trained conditions, or do they generalize, to other tasks or stimuli? This question finds its significance in theoretical as well as practical implications. From an applied perspective, the evaluation of generalization of learning in different auditory perceptual skills can guide the design of training protocols used to improve hearing abilities necessary for certain tasks. For a recent review on generalization of auditory learning see (Wright and Zhang 2009).

The present study examined learning and generalization using discrimination of sinusoidal amplitude modulation rate (SAM-rate) and sinusoidal frequency modulation rate (SFM-rate). These stimuli were selected because amplitude- and frequency-modulated sounds are the basis of many complex sounds, and thus they represent good candidates to be used in training systems aiming at improving the acquisition of auditory skills; or skills in other modalities that can benefit from auditory information. The motivation of this work originated from research conducted under the framework of the European Project SKILLS (www.skills-ip.eu). The major goal of this project is to develope training protocols to improve transfer of skills using novel multimodal and virtual reality technologies. A particular scenario within this project is concerned with the development of a surgical simulator for training on a surgery that involves difficult and critical drilling procedures. The reason for considering the auditory modality is that it has been reported that expert surgeons do extract information from drilling sounds to fine tune forces and directions whereas novice trainees do not (Praamsma et al. 2008).

As mentioned above, it is also of interest to evaluate the generalization of learning. The literature on auditory learning and generalization of learning on discrimination of rate of modulation is scarce. The study by Grimault et al. (2003) examined the hypothesis that pitch of unresolved harmonics is based on the estimation of amplitude modulation rate in the envelope of the sound. Two groups trained on AM-rate discrimination using a narrowband noise carrier. One group trained with a nominal base modulation rate of 88 Hz and the other group trained with a nominal base modulation rate of 250 Hz. Both groups showed significant learning and generalization to a fundamental frequency discrimination task. However, training in SAM-rate discrimination using a narrowband noise carrier did not result in significant improvements in fundamental frequency discrimination based on unresolved harmonics as compared to resolved harmonics. In a second study, learning and generalization of SAM-rate discrimination have been investigated using a wideband carrier (Fitzgerald and Wright 2005). Their results showed that learning generalized to higher modulation rates, but it did not generalize to lower rates nor to other tasks such as pure-tone frequency discrimination and SAM detection. To the best of the author's knowledge, there are no studies that address learning of modulation-rate discrimination using sinusoidal carriers.

Along with the study of auditory learning *per se*, it is important to note that a generalization-of-learning paradigm can shed some lights about the mechanisms underlying perception of amplitude and frequency modulation rate. That is, if the same mechanism underlies perception of both amplitude and frequency modulation rate, one would expect to observe transfer of learning between the perception of the two types of modulation. For example, using harmonic complex tones in a fundamental frequency discrimination task, generalization of learning has been used to support the hypothesis that two different mechanisms mediate the encoding of pitch depending on whether the harmonics are resolved or unresolved (Grimault et al. 2002).

METHOD

Stimuli

Stimuli consisted of 120-ms modulated tones with a nominal carrier frequencies of 800 Hz, and base modulation rates of 32 Hz or 128 Hz. All stimuli included 20-ms of raised-cosine on-off gatings, and were presented with an inter-stimulus interval of 200 ms. For each trial the initial modulation phase was selected randomly from the interval $[0, 2\pi)$, and the initial carrier phase was always zero. Modulation depth was constant and at suprathreshold for both modulation types. For SAM stimuli the modulation depth was fixed to 80%. For SFM stimuli the modulation depth Δf was fixed to 10% the frequency of the nominal carrier. Stimuli were digitally generated by custom-made software written in Python and running on a Linux PC. The PC was equipped with a professional audio card (RME Hammerfall DSP 9632) whose digital output was connected to a 24-bit D/A converter (Swisssonic DR24) set at a sampling rate of 48 kHz. From the D/A converter the signal was fed to an amplifier (Rotel RB 976 mkII). To reduce the amplifier's noise floor a custom-made 20-dB passive attenuator was connected to the output of the amplifier. The output signal from the attenuator was delivered to the listeners over the left earpiece of equalized Beyerdynamic DT-990 circumaural headphones. The overall level of the unmodulated carrier was 68 dB SPL as measured in a head and torso simulator (B&K Type 4128).

Procedure and conditions

The experimental protocol consisted of a pre-test, training phase, and post-test. listeners were divided into two groups of equal size (n=3). Both groups completed all phases. On the pre- and post-tests, discrimination thresholds were measured for four conditions formed by the combination of base modulation rate (32 and 128 Hz) and modulation type (SAM and SFM). The four conditions were presented in a random order to each listener, but the same order was used for the pre- and post-tests. Before the pre-test all listeners performed a familiarization session in which thresholds for all conditions were measured at least one time. For the pre- and post-tests four thresholds were measured for each condition, and all measurements were completed within one session of 2 hours. During training, the base modulation rate was fixed to 32 Hz, and one group trained on SAM-rate discrimination and the other group on SFM-rate discrimination. Each listener completed eight 1-h sessions, which were administered during a period of 3 to 4 weeks with 2 to 3 sessions per week and only one session allowed per day. Eighteen thresholds were measured on each session. The mean time between pre-test and post-test was 33.6 days (sd=2.9). The complete experiment was carried out with listeners seated individually in a double-walled sound attenuating room.

Discrimination of SAM- and SFM-rate was measured using an adaptive three-interval three-alternative forced-choice procedure (3I-3AFC). listeners were presented with three consecutive sound stimuli, two standard stimuli having a base modulation rate of f_m and one test stimulus having a rate of $f_m + \Delta f_m$. The order of the presentation was randomized. Listeners were instructed to identify the interval containing the different sound, i.e. the test stimulus. After two consecutive correct responses the difference between standard and test stimuli was reduced, and after one incorrect response this difference was increased. This rule targets the 70.7% correct point in the psychometric function (Levitt 1971). The initial Δf_m was set equal to the base modulation rate and then it was increased or decreased by a factor of 2 until the fourth reversal and by $\sqrt{2}$ thereafter. The adaptive track stopped after 12 reversals and the threshold was computed from the geometric average of the last 8 reversals. To minimize the possibility that listeners might follow the progression of the adaptive track, two tracks were randomly interleaved, i.e. two thresholds were measured simultaneously. Listeners responded by pressing one of three buttons in a custom made button-box connected to the PC via the parallel port. This box had lights above the buttons that were used to visually cue the interval being presented. Only during training, the lights were also used to provide correct-answer feedback to the listeners. On a trial-by-trial basis, the frequency of the carrier was roved within an interval equal to 10% the nominal carrier and centered around the carrier. This was done to increase the level of difficulty (potential for more room for improvement), and to discourage listeners from using a long-term memory of the standard stimulus and focus their attention on comparing the three intervals within a trial.

Listeners

Six paid listeners (2 females and 4 males) were recruited from the student population of Aalborg University. Their ages ranged from 22 to 24. There were two males and one female assigned to each training group. All listeners had absolute thresholds less than 20 dB hearing level at all audiometric frequencies (250-8000 Hz in octave steps), and at the frequency of the sinusoidal carrier employed in this study. Listeners had no previous experience on psychoacoustic experiments.

RESULTS

Learning on trained conditions

The results for the individual listeners are plotted in Figure 1. Listeners S1, S2, and S3 trained on SAM-rate discrimination, and listeners S4, S5, S6 trained on SFM-rate discrimination.

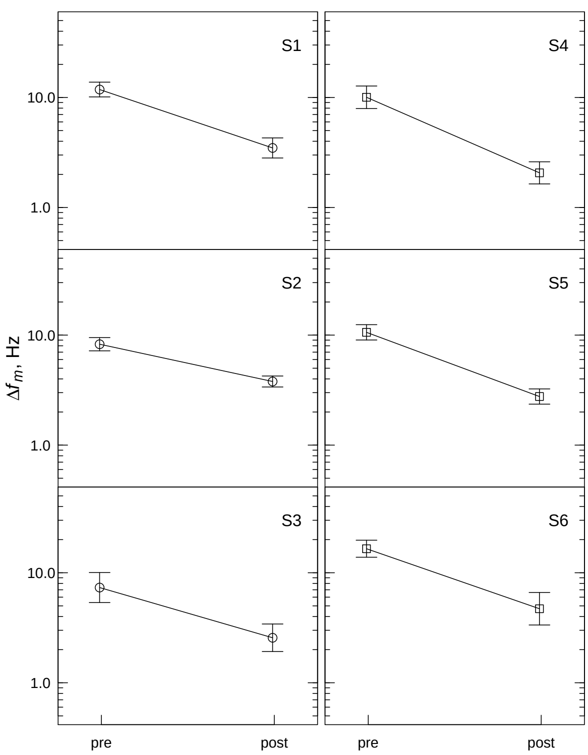
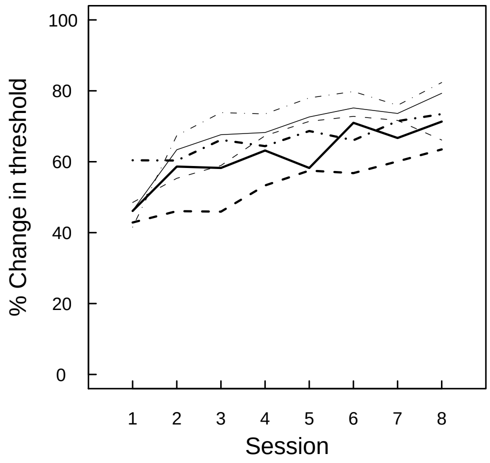
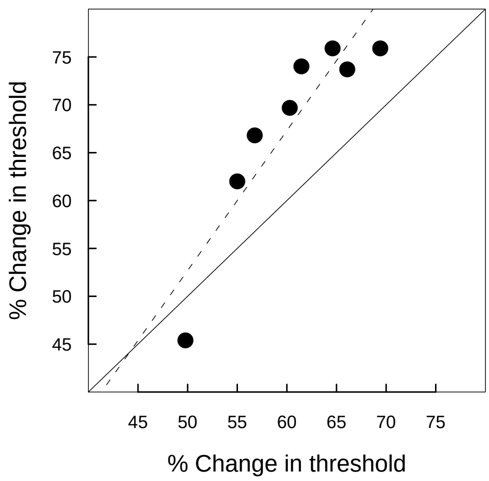


Figure 1: Learning on trained conditions. Individual thresholds of the six listeners are shown on the different panels. Listeners S1, S2, and S3 trained on SAM-rate discrimination, and listeners S4, S5, and S6 trained on SFM-rate discrimination. For both conditions the carrier frequency and base modulation rate were fixed to 800 Hz and 32 Hz respectively. Error bars indicate ± one standard deviation.



(a)



(b)

Figure 2: (a) Normalized percentage change thresholds for each listener as a function of training session. Thick lines represent training on SAM-rate discrimination and thin lines represent training on SFM-rate discrimination. (b) Mean normalized percentage change thresholds for SAM-rate (x-axis) and for SFM-rate (y-axis) are shown for the different training sessions. The regression line is shown by the dashed line. The solid diagonal indicates a regression-line slope of 1.

All listeners showed a clear improvement in performance. Results were pooled across listeners from the same training group and two-sample t-tests were performed on the log-transformed thresholds. This analysis indicated that post-test thresholds were significantly lower than pre-test thresholds for both the SAM-rate group (t(22)=9.04, p<0.0001) and the SFM-rate group (t(22)=9.37, p<0.0001). The improvement in performance was slightly better for the SFM-rate group because pre-test thresholds for this group were just significantly larger than those of the SAM-rate group (t(22)=-2.53, p=0.02), and post-test thresholds between the two groups were not significantly different (t(22)=0.53, p=0.603).

To examine the progression of learning during training we calculated the normalized percentage change in threshold ($\%nTH$) relative to the pre-test thresholds for each listener similar to the analysis peformed by Kumpik et al. (2009). This normalized threshold was calculated as follows

$$\%nTH = \frac{(threshold_{pre-test} - threshold_{sessions})}{threshold_{pre-test}} \times 100 \quad (1)$$

Figure 2a shows the individual learning curves as characterized by the normalized percentage change. A curve with positive slope indicates potential learning. There was a large initial improvement between pre-test and the first training session in a range of approximately 40% to 60% across listeners. This is most probably caused by the integration of feedback in the task. By the end of the training period listeners improved their thresholds within a range of 58% to 80% relative to their pre-test thresholds. To better identify individual listeners as having learned across the eight sessions, linear models were fitted to the individual curves, and one-way analysis of variance (ANOVA) was performed on the log-transformed thresholds with session as factor (Zhang and Wright 2009). If the results from these analyzes were significant then the listeners were considered as having learned. All listeners met the criteria, i.e. all listeners significantly improved through training (ANOVA: all p<0.01, linear regression: all slopes significantly greater than zero, all p<0.05). To analyze differences in learning rates between training groups, a linear regression was fitted to the mean normalized percentage changes. Figure 2b shows the mean normalized percentage change thresholds for the SAM-rate group (x-axis) plotted against the mean normalized percentage change thresholds for the SFM-rate group (y-axis). Each point corresponds to data from one session. The result from the regression analysis ($y = -20.387 + 1.462x$) revealed that on average listeners from the SFM-rate group learned faster than listeners from the SAM-rate group. All but one data point were above the solid diagonal line.

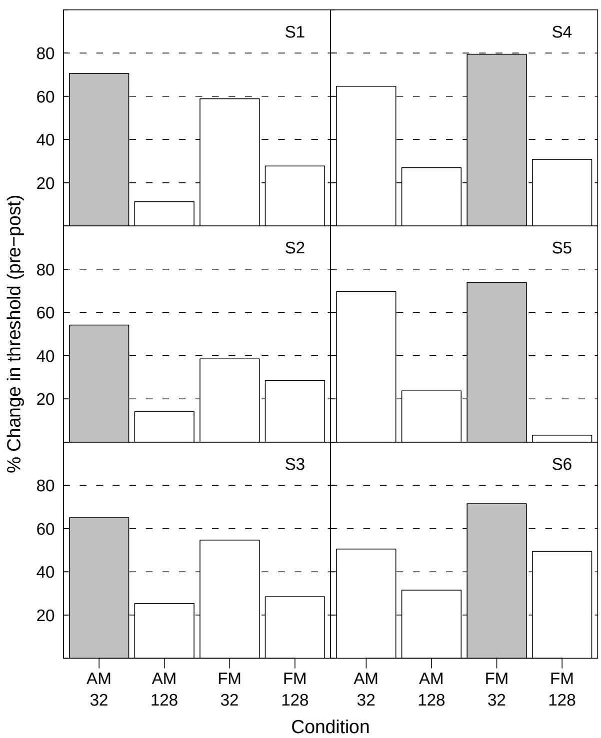


Figure 3: Normalized percentage change thresholds for each listener between pre- and post-tests. Grey bars correspond to the trained condition.

.

Generalization to untrained conditions

Normalized percentage change thresholds for all conditions are shown in Figure 3. Post-test thresholds for all conditions and listeners were lower than their corresponding pre-test thresholds, thus indicating potential transfer of learning to untrained conditions. The largest improvement in performance was observed on the trained condition (grey bars), and the amount of transfer was larger across modulation type than across modulation rate. Two-sample t-test analysis comparing pre- and post-test log-transformed thresholds pooled across subjects revealed that for both training groups learning generalized across modulation type (SAM-to-SFM, $t(22)=-5.10$, $p<0.0001$; SFM-to-SAM, $t(22)=-7.03$, $p<0.0001$). The amount of generalization across modulation type for the 32-Hz modulation rate was comparable between the two training groups. Although post-test thresholds were lower than pre-test thresholds for the 128-Hz modulation rate, differences did not reach significance for any of the two training groups.

DISCUSSION AND CONCLUSION

Modulation rate spans three sensory categories. At low rates (approx. < 15 Hz) loudness and pitch fluctuations are perceived for AM and FM sounds respectively. As the rate increases beyond about 20 Hz the fluctuations gradually become constant and a sensation of "roughness" appears. As the rate increases even more, roughness decreases giving place to a rising sensation of pitch (Moore 2003). The range of modulation rate (32–64 Hz) and the carrier frequency (800 Hz) to which listeners were exposed during training corresponds to the perceptual range in which the hearing sensation of roughness is assumed to operate (Zwicker and Fastl 1990). That is, listeners improved their performance at discriminating which sound was rougher or had a faster vibration. The observation that greatest improvement was on the trained condition, and that greatest generalization was across modulation type, for the same modulation rate, is in good agreement with the understanding that roughness is associated to rapid variations of either amplitude or frequency.

The present data showed partial generalization of learning across modulation rate. Athough improvements in the untrained modulation rate was not significant, it is still important to discuss this generalization in connection to previous studies. Fitzgerald and Wright (2005) trained 9 listeners on SAM-rate discrimination using a base modulation rate of 150 Hz and a wideband Gaussian noise carrier. Listeners completed 720 trials per day during a period of 6 to 8 days. Fitzgerald and Wright reported that learning generalized partially to a higher modulation rate (300 Hz), but not to a lower modulation rate (30 Hz). Using a narrowband noise carrier (1375-1875 Hz) Grimault et al. (2003) also reported generalization of learning between 88 and 250 Hz of SAM rate. The present results are in agreement with partial generalization of modulation-rate discrimination to a higher modulation rate. However, comparison between results from these studies and the present one should be treated with care. The reason is that the major argument for their generalization is based on that SAM-rate discrimination for 150 Hz and 300 Hz is mediated by pitch cues, but SAM-rate discrimination for 30 Hz is not. According to the perceptual range of modulation rate, SAM-rate discrimination for 30 Hz is likely to be based on roughness perception.

Note that for a carrier frequency of 800 Hz the range of modulation rate between 32 and 128 Hz may include the transition between roughness and pitch perception. By using the equivalent rectangular bandwidth (Moore 2003) one can calculate the bandwidth of the auditory filter centered at 800 Hz (approx. 110 Hz or $\pm$55 Hz from the center frequency). Thus, for a modulation rate of 128 Hz or higher the sideband components lie outside the critical bandwidth, and the individual sinusoidal components are said to be resolved resulting in pitch perception. The assumption that generalization occurs across stimuli that elicit the same percept does not find support from this analysis. The partial generalization of learning to a higher modulation rate may be explained by assuming that there is certain overlapping between the sensory categories spanned by modulation rate. If the range of modulation rate 32–128 Hz is not large enough to provide a clear separation between roughness- and pitch-based cues, then, it is possible that discrimination for a base modulation rate of 128 Hz may still retain some roughness sensation that listeners were able to capitalize on because they have learned it. Another possible explanation is that the observed partial generalization does not reflect pure perceptual learning but also task-related or procedural learning. That is, learning may have resulted more from an improvement in storing and comparing standard stimuli and test stimuli. This explanation has been previously suggested by Karmarkar and Buonomano (2003) in their study on learning using a temporal auditory task.

Although conclusions on these findings should be considered as preliminary, if a protocol shall be designed to improve roughness perception for instead, training on FM-rate discrimination may be given priority over training on AM-rate discrimination. From an applied perspective particular to the surgical simulator mentioned in the introduction, these findings may be relevant if one considers that the sound of surgical drilling is very likely to have a strong "roughness" component. If one wants to boost performance on modulation-rate discrimination for both SAM- and SFM-rate discrimination it appears that a training protocol based on frequency modulation is better than a protocol based on amplitude modulation. This is derived from the observation that the SFM-rate group showed more learning specific to the trained condition than the SAM-rate group, and also a greater generalization to the untrained modulation type. In addition, the rate of learning was faster for the SFM-rate group than for the SAM-rate group.

A noteworthy aspect of these results is the sudden improvement observed between the pre-test phase and the first training session. This sudden improvement is attributed to the use of correct-response feedback during training. This accelerated learning effect may be explained by the view that feedback can enhance motivation, which in turn can positively modulate auditory perceptual learning (Amitay et al. 2010). Albeit the effect of feedback was not directly addressed in this work, support to this view was found from informal reports from some listeners describing that feedback immediately increased their level of confidence when performing the task.

In summary, multiple training sessions can improve performance on SAM- and SFM-rate discrimination, and learning can generalize across modulation type but weakly to modulation rate. The fact that carriers differ across the revised studies and the present one, together with the observation that learning partially generalizes to a higher rate, suggests that generalization across modulation rate may be independent of stimulus carrier.

ACKNOWLEDGMENTS

This work was supported by the SKILLS Integrated Project (IST-FP6 #035005, http://www.skills-ip.eu) funded by the European Commission.

REFERENCES

Amitay, Sigal, David J. C. Hawkey, and David R. Moore (2005). "Auditory frequency discrimination learning is affected by stimulus variability". *Perception & Psychophysics* 67.4, pp. 691–698.

Amitay, Sygal et al. (2010). "Motivation and Intelligence Drive Auditory Perceptual Learning". *PloS one* 5.3, e9816.

Demany, Laurent (1985). "Perceptual learning in frequency discrimination". *J. Acoust. Soc. Am.* 78.3, pp. 1118–1120.

Fitzgerald, Matthew B. and Beverly A. Wright (2005). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise". *J. Acoust. Soc. Am.* 118, pp. 3794–3803.

Grimault, Nicolas et al. (2002). "Evidence for two pitch encoding mechanisms using a selective auditory training paradigm". *Perc. Psychophys.* 64, pp. 189–197.

Grimault, Nicolas et al. (2003). "Learning in discrimination of frequency or modulation rate: generalization to fundamental frequency discrimination". *Hearing Research* 184, pp. 41–50.

Karmarkar, Uma R. and Dean V. Buonomano (2003). "Temporal Specificity of Perceptual Learning in an Auditory Discrimination Task". *Learning Memory* 10, pp. 141–147.

Kumpik, Daniel et al. (2009). "Specificity of binaural perceptual learning for amplitude modulated tones: A comparison of two training methods". *J. Acoust. Soc. Am.* 125.4, pp. 2221–2232.

Levitt, H. (1971). "Transformed up-down methods in psychophysics". *J. Acoust. Soc. Am.* 49, pp. 467–477.

Lu, Zhong-Lin et al. (2009). "Perceptual learning: Functions, mechanisms and applications". *Vision Research* 49.21, pp. 2531–2534.

Moore, Brian C. J. (2003). *An Introduction to the Psychology of Hearing*. 5th. Academic Press.

Praamsma, Monate et al. (2008). "Drilling sounds are used by surgeons and intermediate residents, but not novice orthopedic trainees, to guide drilling motions". *Can J Surg* 51.6, pp. 442–446.

Wright, Beverly A. and Yuxuan Zhang (2009). "A review of the generalization of auditory learning". *Phil. Trans. R. Soc. B* 364, pp. 301–311.

Zhang, Ying, Adrian R. L. Travis, and Neil Collings (2007). "Evaluation of Multi-sensory Feedback on the Usability of a Virtual Assembly Environment". *Journal of Multimedia* 2.1, pp. 38–47.

Zhang, Yuxuan and Beverly A. Wright (2009). "An influence of amplitude modulation on interaural level difference processing suggested by learning patterns of human adults". *J. Acoust. Soc. Am.* 126.3, pp. 1349–1358.

Zwicker, Eberhard and Hugo Fastl (1990). *Psychoacoustic: Facts and Models*. Berlin, Germany: Springer-Verlag.